Vol. 10, No. 1, April 2025, Page. 120-129
https://journal.uin-alauddin.ac.id/index.php/instek
p-ISSN: 2541-1179
e-ISSN: 2581-1711
JURNAL INSTEK
INFORMATIKA SAINS DAN TEKNOLOGI

PREDIKSI STATUS ANEMIA DENGAN PENDEKATAN PEMBELAJARAN MESIN ALGORITMA *SUPPORT VECTOR MACHINE* (SVM) DAN SELEKSI FITUR *FIREFLY ALGORITHM*

Tiara Meylinda Sari[*1], Lukman Hakim[2]

[1,2]Program Studi Teknik Informatika, Universitas Yudharta Pasuruan
Email: [1]tiarameylindasari@gmail.com, [2]lukman@yudharta.ac.id

Abstrak

Anemia merupakan kelainan tubuh yang ditandai dengan rendahnya kadar hemoglobin (Hb) dalam sel darah, dan dapat menjadi masalah kesehatan yang serius terlebih pada remaja perempuan jika tidak segera diobati dengan baik. Penelitian ini dilakukan untuk memprediksi status anemia berdasarkan data pasien menggunakan algoritma SVM dengan pemilihan fitur *Firefly Algorithm* untuk meningkatkan akurasi. Pengujian dilakukan dengan menggunakan empat *kernel* algoritma SVM yaitu *kernel Linear, Polynomial*, RBF dan *Sigmoid*. Hasil penelitian menunjukkan bahwa penggabungan FA dan SVM dapat meningkatkan akurasi pada tiga kernel SVM yaitu *kernel linear* dari akurasi awal 98.95% menjadi 99.65%, *kernel polynomial* akurasi awal 94.74% menjadi 98.60%, pada *kernel* RBF akurasi awal 93.68% menjadi 98.25%, namun pada *kernel sigmoid* akurasi mengalami penurunan dari akurasi awal 47.02% menjadi 12.98%. Kesimpulannya, penerapan FA untuk memilih fitur-fitur penting pada SVM efektif dan berdampak pada peningkatan akurasi untuk tiga kernel SVM dan penurunan akurasi pada satu kernel hal tersebut terjadi karena *underfitting*. Seleksi fitur menjadi efektif jika menghasilkan kombinasi fitur yang tepat dan dapat menjadi tidak efektif jika menghasilkan kombinasi fitur yang tidak tepat.

Kata kunci: Akurasi, Anemia, *Firefly Algorithm*, Seleksi Fitur, SVM.

Abstract

Anemia is a body disorder characterized by low levels of hemoglobin (Hb) in blood cells, and can be a serious health problem, especially in adolescent girls if not treated properly. This study was conducted to predict anemia status based on patient data using the SVM algorithm with the selection of Firefly Algorithm features to improve accuracy. Testing was carried out using four SVM algorithm kernels, namely Linear, Polynomial, RBF and Sigmoid kernels. The results showed that the combination of FA and SVM can improve accuracy in three SVM kernels, namely the linear kernel from an initial accuracy of 98.95% to 99.65%, the polynomial kernel initial accuracy of 94.74% to 98.60%, in the RBF kernel initial accuracy of 93.68% to 98.25%, but in the sigmoid kernel the accuracy decreased from an initial accuracy of 47.02% to 12.98%. In conclusion, the application of FA to select important features in SVM is effective and has an impact on increasing accuracy for three SVM kernels and decreasing accuracy in one kernel, this occurs due to underfitting. Feature selection is effective if it produces the right combination of features and can be ineffective if it produces the wrong combination of features.

Keywords: *Accuracy, Anemia, Feature Selection, Firefly Algorithm, SVM.*

This work is an open access article and licensed under a Creative Commons Attribution-NonCommercial ShareAlike 4.0 International (CC BY-NC-SA 4.0)

1. PENDAHULUAN

Anemia merupakan penyakit umum yang dapat terjadi di masyarakat [1]. Anemia yaitu kondisi ketika tubuh kekurangan hemoglobin (Hb) pada sel darah merah atau kadar Hb kurang dari 12 g/dL pada perempuan serta kurang dari 13,5 g/dL pada laki-laki [2]. Anemia merupakan masalah kesehatan yang dapat terjadi pada anak-anak, remaja, ibu hamil, bahkan lansia [3]–[5]. Anemia dapat menyebabkan pertumbuhan terhambat, gangguan kecerdasan dan motorik, penurunan kekebalan tubuh serta penurunan kesuburan [6]. Potensi anemia terhadap perempuan lebih tinggi 6% daripada laki-laki. Bagi remaja putri, anemia mempunyai pengaruh berkelanjutan seperti kelahiran prematur, risiko kematian yang meningkat saat melahirkan, berat bayi lahir rendah (BBLR), dan terhambatnya pertumbuhan anak (stunting). Berdasarkan Riset Kesehatan Dasar (Riskedas) Indonesia tahun 2018, sebanyak 26,8% anak usia 5 hingga 14 tahun dan 32% anak usia 15 hingga 24 tahun menderita anemia,

atau 3 hingga 4 dari 10 orang di sekitar kita memiliki anemia. WHO memperkirakan di seluruh dunia penderita anemia sebanyak 40% anak usia 6 hingga 59 bulan, 37% wanita hamil dan 30% wanita berusia 15 hingga 49 tahun. Jika gejala anemia tidak segera diketahui dan diobati dengan segera, hal ini dapat menimbulkan dampak buruk bagi penderitanya. Keakuratan dalam prediksi sebuah penyakit sangat penting dan memerlukan keputusan yang tepat serta efektif untuk mengambil suatu analisa. Oleh karena itu, diperlukan penyelesaian dengan teknologi database yang dapat membantu mengklasifikasikan keadaan penyakit secara otomatis berdasarkan data penyakit pasien.

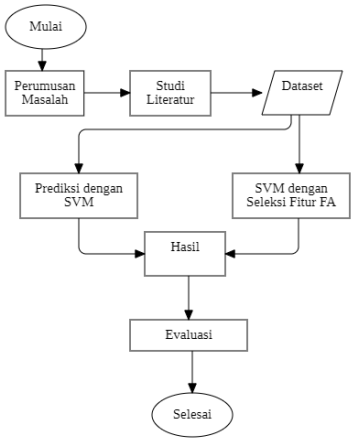
Kondisi dataset anemia di Indonesia masih kurang terstandarisasi karena seringkali dikumpulkan dari berbagai sumber seperti puskesmas atau rumah sakit dengan standar yang berbeda-beda sehingga dapat menyebabkan banyak fitur tidak relevan yang mempengaruhi proses prediksi juga beberapa fitur dapat menjadi lebih dominan, sehingga diperlukan penambangan data. Penambangan data adalah teknik untuk membentuk model pembelajaran mesin [7]. Prediksi *data mining* menggunakan model atau algoritma matematika untuk memperkirakan nilai atau hasil masa depan [8] berdasarkan data historis. Proses pengambilan keputusan berdasarkan data sepenuhnya selaras dengan upaya transformasi digital [9]. Analisis data historis dilakukan untuk mendapatkan pola atau hubungan yang penting [10]. Berdasarkan pola-pola ini, model prediktif seperti regresi, pohon keputusan, dan jaringan saraf tiruan dibuat. Setelah model dilatih berdasarkan data historis, model tersebut dapat digunakan untuk memprediksi hasil berdasarkan data baru. Beberapa model yang sering digunakan dalam penambangan data yaitu, Naive Bayes Classifier, Decision Tree, Support Vector Machine (SVM), Random Forest dan K-Nearest Neighbor (KNN) [11]. Metode-metode tersebut memiliki kelebihan dan kekurangan masing-masing. Salah satu kelebihan dari metode klasifikasi tersebut dapat dilihat berdasarkan cepat lambatnya metode ini memproses objek dataset hingga mendapatkan hasil akurasi yang tinggi dalam proses prediksi dan klasifikasi. Namun, beberapa metode mungkin memiliki performa yang buruk saat memproses kumpulan data yang sangat besar, sehingga menyebabkan nilai akurasi yang lebih rendah.

Algoritma Support Vector Machine (SVM) dapat menunjukkan hasil klasifikasi yang baik [12], [13]. Meskipun SVM efektif untuk data besar, waktu pelatihan SVM bisa lebih lambat untuk kumpulan data dengan banyak fitur dengan data yang sangat besar. Dengan dilakukan seleksi fitur dapat mengurangi jumlah fitur dalam data yang akan diproses. Dataset yang digunakan dalam penelitian ini memiliki 6 atribut, meskipun jumlah fitur dalam dataset sedikit seleksi fitur masih tetap diperlukan karena dengan seleksi fitur dapat meningkatkan akurasi model pembelajaran, dimana beberapa fitur dapat bersifat noise atau tidak relevan sehingga dapat menurunkan performa model pembelajaran, dan dengan lebih sedikit fitur model akan menjadi lebih mudah memahami pola serta pelatihan dan prediksi model menjadi lebih efisien. Oleh karena itu, pemilihan fitur dapat digunakan untuk mengoptimalkan algoritma SVM dengan menggunakan algoritma *firefly* (FA) untuk meningkatkan akurasi dan mempercepat prediksi [14]. Seleksi fitur merupakan proses pemilihan subset dari fitur relevan [15] yang paling berpengaruh dalam model pembelajaran mesin. *Support Vector Machine* (SVM) adalah metode klasifikasi vektorisasi instruksional yang diperkenalkan oleh Vapnik pada tahun 1995 [16]. *Support Vector Machine* (SVM) telah banyak digunakan untuk mengatasi masalah *linear* juga *nonlinear* [17]. Dalam penelitian [11], [18], [19] SVM dapat memberikan akurasi yang baik dalam prediksi penyakit. *Firefly Algorithm* merupakan sebuah algoritma metaheuristik yang terinspirasi oleh perilaku kunang-kunang dalam menarik pasangan melalui sinyal cahaya [20]–[22]. Penggunaan *Firefly Algorithm* (FA) dalam prediksi suatu penyakit sangat jarang dilakukan akan tetapi dalam hal lain seperti pada penelitian [23], [24] *Firefly Algorithm* (FA) dapat meningkatkan akurasi model SVM. Hasil penelitian yang dilakukan oleh Styawati menunjukkan bahwa algoritma *Firefly* mampu mendapatkan kombinasi parameter SVM yang sesuai [24]. Penelitian yang dilakukan oleh Pujianto et al., menunjukkan ketepatan dalam penentuan kelas sentimen menggunakan metode *Support Vector Machine* (SVM) sebesar 85%, setelah dioptimasi menggunakan algortima *Firefly* menghasilkan ketepatan sebesar 86% [23]. Algoritma SVM sering digunakan dalam prediksi penyakit namun penggabungan SVM dengan Firefly masih jarang di temukan untuk prediksi penyakit anemia. Pada penelitian ini bertujuan memberikan solusi untuk meningkatkan akurasi dalam memprediksi penyakit anemia berdasarkan data anemia. Prediksi dilakukan menggunakan algoritma *Support Vector Machine* (SVM) dengan seleksi fitur *Firefly Algorithm* (FA) untuk meningkatkan akurasi prediksi. Akurasi adalah tingkat ketelitian dalam memprediksi suatu data yang benar dan salah [25]. Algoritma *firefly* dapat digunakan untuk mengatasi masalah pemilihan fitur [26] dan dapat meningkatkan efisiensi klasifikasi [27]–[30].

Vol. 10, No. 1, April 2025, Page. 120-129
https://journal.uin-alauddin.ac.id/index.php/instek
p-ISSN: 2541-1179
e-ISSN: 2581-1711
JURNAL INSTEK
INFORMATIKA SAINS DAN TEKNOLOGI

2. METODE PENELITIAN

Metode penelitian sangat diperlukan untuk mencapai hasil yang selaras dengan tahapan yang sudah ditetapkan. Tahap penelitian ditampilkan pada Gambar 1**.**



Gambar 1. Alur Penelitian

2.1. Perumusan Masalah

Anemia adalah masalah kesehatan internasional yang serius di Masyarakat, terutama dapat menyerang anak-anak, perempuan remaja, perempuan yang mengalami haid/menstruasi, serta ibu hamil juga saat nifas. Apabila gejala anemia tidak segera disadari dan ditangani dapat berakibat fatal bagi penderitanya. Oleh karena itu solusi dari teknologi basis data diperlukan yang dapat secara otomatis akan membantu mengklasifikasikan dan memprediksi keadaan penyakit berdasarkan data.

2.2. Studi Literatur

Pada penelitian ini referensi dicari melalui internet, artikel ilmiah, laporan penelitian, skripsi, jurnal dan hasil penelitian sebelumnya. Tujuannya adalah untuk memperkuat masalah dan memperkuatnya sebagai alasan serta dasar teori dalam melakukan desain penelitian.

2.3. Dataset

Dataset dalam penelitian didapatkan dari situs online Kaggle (https://www.kaggle.com/datasets/biswaranjanrao/anemia-dataset) dalam bentuk file csv. Setelah data diambil dilakukan pengecekan data untuk mengetahui kondisi data dan didapatkan informasi bahwa tidak ada data yang hilang/kosong sehingga tidak dilakukan *data cleaning* juga semua tipe data sudah sesuai sehingga tidak ada yang harus diubah dari data aslinya dan dapat langsung digunakan dalam prediksi. Data tersebut terdiri atas 1.421 data dengan lima atribut dan satu sebagai label yaitu *Result* (hasil) dengan 0 menyatakan tidak anemia dan 1 berarti anemia. Terdapat 801 data sebagai tidak anemia dan 620 data sebagai anemia. Berikut tabel 1 atribut-atribut yang akan digunakan yaitu:

Tabel 1. Atribut – atribut Dataset

Atribut	Keterangan	Tipe data	Range
Gender	Kategori *Gender*	Kategori	0 = laki-laki, 1 = Perempuan
Hemoglobin	Parameter dari Hemoglobin	Numerik	6.6 – 16.9
MCH	Parameter dari *Mean Corpuscular Hemoglobin* (MCH)	Numerik	16 – 30
MCHC	Parameter dari *Mean Corpuscular Hemoglobin Concentration* (MCHC)	Numerik	27.8 – 32.5

p-ISSN: 2541-1179
e-ISSN: 2581-1711
JURNAL INSTEK
INFORMATIKA SAINS DAN TEKNOLOGI

MCV	Parameter dari Mean Corpuscular Volume (MCV)	Numerik	69.4 – 101.6
Result	Kategori klasifikasi Anemia atau Tidak Anemia	Kategori	0 = Tidak Anemia, 1 = Anemia

2.4. Prediksi dengan SVM

Pada penelitian ini akan digunakan 4 kernel algoritma SVM. Penggunaan jenis kernel yang berbeda dapat memaksimalkan akurasi model yaitu dengan memberikan informasi yang berbeda berdasarkan hasil akurasi, presisi, F1-*Score* sehingga dapat memberikan pilihan atas performa kernel manakah yang terbaik dan cocok dengan data dibandingkan jika hanya menggunakan satu kernel saja. Berikut ini adalah penjelasan tentang kernel tersebut beserta dengan rumusnya, yaitu:

2.4.1. *Kernel Linear*

Kernel linear merupakan kernel yang paling sederhana. Jika data dapat dipisahkan secara linear maka dapat menggunakan kernel tersebut.

$$K(x, y) = (x \,.\, y) \qquad [31]\,(1)$$

K = nilai kernel, x = data uji, y = data latih.

2.4.2. *Kernel Polynomial*

Penggunaan *kernel polynomial* yaitu ketika hubungan antara kelas dan fitur tidak linear atau untuk data yang dapat dipisahkan secara non-linear.

$$K(x, y) = (x \,.\, y \; + \; c)^d \qquad [31](2)$$

Di mana c adalah konstanta bebas dan d adalah derajat polinomial.

2.4.3. Kernel RBF

Kernel RBF adalah salah satu kernel yang paling populer dan banyak digunakan [32]. Kernel RBF sangat efektif ketika tidak diketahui sebelumnya tentang distribusi data.

$$K(x, y) \; = \; \exp\left(-\gamma \,||\, x \,.\, y \,||^2\right) \qquad [31](3)$$

Di mana γ adalah parameter yang menentukan seberapa jauh pengaruh satu sampel tunggal.

2.4.4. Kernel Sigmoid

Kernel sigmoid mirip dengan fungsi aktivasi dalam jaringan saraf buatan. Kernel ini digunakan dalam beberapa aplikasi seperti jaringan syaraf tiruan.

$$K(x, y) \; = \; \tanh\left(\alpha x \,.\, y \; + \; c\right) \qquad (4)$$

Di mana α dan c adalah parameter yang bisa disesuaikan.

2.5 SVM dengan Seleksi Fitur FA

Setelah dilakukan prediksi menggunakan algoritma SVM saja selanjutnya dilakukan prediksi gabungan antara algoritma SVM dengan FA untuk diketahui perbedaan akurasi. *Firefly Algorithm* merupakan sebuah algoritma metaheuristik yang terinspirasi oleh perilaku kunang-kunang dalam menarik pasangan melalui sinyal cahaya [20]. Diperkenalkan oleh Xin-She Yang pada tahun 2008. Algoritma ini memanfaatkan dua karakteristik utama kunang-kunang: intensitas cahaya yang memudar dengan jarak dan kemampuan kunang-kunang untuk berpindah menuju Cahaya yang lebih terang [33]. Algoritma ini bekerja dengan memperbarui posisi kunang-kunang [21], di mana setiap kunang-kunang akan bergerak mendekati kunang-kunang lain yang memiliki intensitas cahaya lebih terang. Proses seleksi fitur FA melibatkan langkah-langkah berikut:

– Inisialisasi

Menentukan jumlah kunang-kunang (n), jumlah iterasi maksimum (max_iter), parameter alpha, beta, gamma, dan dimensi masalah (d, yang merupakan jumlah total fitur). Dalam penelitian telah ditentukan max_iter = 10 serta alpha, beta, gamma masing-masing 0,1 penentuan ini telah didasarkan pada pertimbangan keseimbangan antara kompleksitas komputasi, waktu pemrosesan dan kualitas solusi.

– Generasi populasi awal

Inisialisasi populasi kunang-kunang secara acak. Setiap kunang-kunang adalah vektor biner dengan panjang d.

– Evaluasi Fitness

Evaluasi fitness dari setiap kunang-kunang berdasarkan subset fitur yang dipilih. Fitness biasanya dihitung berdasarkan kinerja model machine learning (misalnya, akurasi, F1-score) yang dilatih menggunakan subset fitur tersebut. Untuk mendapatkan nilai fitness dilakukan dengan perhitungan pada persamaan-persamaan berikut. Kunang-kunang *i*, intensitas cahaya I_i bisa dihitung sebagai:

$$I_i = f(x_i) \quad (5)$$

I_i = kecerahan cahaya *firefly* ke-i, *f* (x) = fungsi objektif .
Jarak r_{ij} antara dua kunang-kunang *i* dan *j* [14]dalam ruang Euclidean dapat dihitung dengan:

$$r_{ij} = || x_i - x_j || = \sqrt{\sum_{k=1}^{d} (x_{i,k} - x_{j,k})^2} \quad (6)$$

r_{ij} = jarak antara *firefly* ke-i dan *firefly* ke-*j*, *d* = jumlah dimensi (fitur), $x_{i,k}$ komponen ke-*k* dari posisi *firefly* ke-*i*.
Daya tarik β dari kunang-kunang *j* terhadap kunang-kunang *i* [14]dapat dihitung sebagai:

$$\beta (r) = \beta_0 e^{-\gamma r2} \quad (7)$$

β_0 = daya tarik maksimum antara 0 dan 1, γ = koefisien serapan cahaya, *r* = jarak *firefly*.

Posisi kunang-kunang *i* diperbarui dengan bergerak ke arah kunang-kunang *j* yang lebih menarik (memiliki intensitas cahaya lebih tinggi) [14] menggunakan rumus:

$$x_i^{t+1} = x_i^t + \beta_0 e^{-\gamma r2} ij (x_j^t - x_i^t) + \alpha (rand - 0.5) \quad (8)$$

$\beta_0 e^{-\gamma r2} ij (x_j^t - x_i^t)$ = Gerakan menuju *firefly* yang lebih terang, α = parameter acak, $rand$ = vektor acak.

– Pembaruan Posisi

Perbarui posisi kunang-kunang berdasarkan intensitas cahaya dan daya tarik, menggunakan rumus-rumus yang dimodifikasi untuk konteks vektor biner.

2.6 Evaluasi

Kinerja model dievaluasi berdasarkan prediksi yang dibuat pada testing set dengan metode *percentage split* yaitu dengan pembagian 20% menjadi data uji dan sisanya yaitu 80% dari data menjadi data latih. Pengetesan akurasi dilakukan dengan menggunakan matriks kebingungan *(confusion matrix). Confusion matrix* adalah matriks yang berfungsi menampilkan gambaran kinerja algoritma dengan membandingkan hasil prediksi model dengan label aktual dari set data uji. Hasil akurasi adalah prediksi (positif dan negatif) yang benar di seluruh data. Ukuran evaluasi kinerja yang digunakan adalah Akurasi, *Precision, Recall* dan F1-*score*.

2.6.1 Akurasi

Akurasi adalah bilangan yang digunakan dalam memperkirakan data benar dan salah [25]. Adapun rumus yang digunakan untuk mengukur akurasi sebagai berikut:

$$\frac{TP+TN}{TP+TN+FP+FN} \quad (9)$$

TP = True Positive, TN = True Negative, FP = False Positive, FN = False Negative.

2.6.2 *Precision*

Presisi merupakan persentasi prediksi positif yang sebenarnya dibandingkan dengan hasil prediksi benar (positif) secara keseluruhan. Adapun rumus yang digunakan sebagai berikut :

$$\frac{TP}{TP+FP} \quad (10)$$

TP = True Positive, FP = False Positive.

2.6.3 *Recall*

Recall adalah perbandingan prediksi positif sebenarnya dengan seluruh data positif sebenarnya [34]. Adapun rumus yang digunakan sebagai berikut:

$$\frac{TP}{TP+FN} \quad (11)$$

TP = True Positive, FN = False Negative.

2.6.4 F1-*Score*

F1-*Score* adalah perbandingan rata-rata *precision* dan *recall* serta memberikan keseimpbangan antara keduanya.

$$2 \text{ X } \frac{Precision \text{ X } Recall}{Precision + Recall} \quad (12)$$

3. HASIL DAN PEMBAHASAN

3.1. Pengujian Data dengan SVM

Penggunaan beberapa kernel dalam penelitian yaitu kernel *Linear*, *Polynomial*, RBF *(Radial Basis Function)* dan Sigmoid. Guna mendapatkan nilai akurasi, *presisi, recall* dan F1-*score* dalam prediksi data anemia dijalankan dengan perhitungan *confusion matrix* menggunakan metode pelatihan (training) *percentage split*. Adapun pembagian *percentage split* yang digunakan yaitu 20% menjadi data uji dan sisanya yaitu 80% dari data menjadi data latih. Berikut hasil prediksi menggunakan algoritma SVM. Lihat tabel 2 untuk hasil perhitungan.

Tabel 2. Hasil Perhitungan

Kernel	Akurasi	Presisi	Recall	F1-Score
Linear	98.95%	97.71%	100%	98.84%
Poly	94.74%	93.80%	94.53%	94.16%
RBF	93.68%	92.97%	92.97%	92.97%
Sigmoid	47.02%	40.17%	36.72%	38.37%

Pada Tabel 2. menunjukan bahwa, SVM dengan kernel *Linear* memiliki nilai akurasi, *precision, recall* dan f1-*score* yang paling tinggi. Selanjutnya kernel *Polynomial* pada urutan ke dua, RBF ke-tiga dan Sigmoid ke-empat dimana akurasi, *precision, recall* dan f1-*score* kernel sigmoid adalah yang paling rendah dan memiliki selisih yang jauh dibandingkan dengan tiga kernel lainnya. Pada saat prediksi dibutuhkan waktu sekitar 8 detik.

3.2. SVM dengan Seleksi Fitur *Firefly Algorithm*

Penggunaan algoritma *Firefly* adalah untuk mengidentifikasi subset fitur yang optimal dari dataset. Algoritma Firefly diimplementasikan untuk mengevaluasi dan memilih fitur-fitur yang paling relevan, dengan tujuan meningkatkan akurasi model prediksi. Berikut hasil fitur yang dipilih berdasarkan *best*

fitness Firefly algorithm dengan parameter *Alpha, Beta, Gamma* masing-masing adalah 0,1. Lihat tabel 3 untuk hasil seleksi fitur.

Tabel 3. Hasil Seleksi Fitur

Percobaan ke-i	N_*firefly*	*Best fitness*	Fitur yang dipilih
1		94.04%	014
2	5	93.68%	01234
3		93.33%	0124
1		95.09%	0123
2	10	95.79%	013
3		98.25%	01
1		98.25%	01
2	15	95.79%	013
3		97.19%	012
1		98.25%	01
2	20	97.19%	012
3		98.25%	01

Setelah dilakukan seleksi fitur menggunakan algoritma *Firefly* didapatkan kombinasi fitur terbaik yaitu fitur ke 0 dan ke 1 dimana fitur tersebut yaitu Gender dan Hemoglobin. Berikut Hasil perhitungan nilai akurasi, *precision*, *recall* dan f1-*score* pada masing masing kernel, dapat dilihat pada tabel 4 berikut:

Tabel 4. Hasil Perhitungan Setelah Seleksi Fitur

Kernel	Akurasi	Presisi	Recall	F1-Score
Linear	99.65%	99.22%	100%	99.61%
Poly	98.60%	96.97%	100%	98.46%
RBF	98.25%	96.24%	100%	98.08%
Sigmoid	12.98%	0%	0%	0%

Setelah dilakukan seleksi fitur SVM dengan kernel *Linear* tetap memiliki nilai akurasi, *precision*, *recall* dan f1-*score* yang paling tinggi. Selanjutnya kernel *Polynomial* pada urutan ke-dua RBF ke-tiga dan Sigmoid ke-empat dimana *precision*, *recall* dan f1-*score* kernel sigmoid tidak ada atau 0% kemungkinan terjadinya seleksi fitur menyebabkan kernel sigmoid mengalami *underfitting* sehingga kernel tidak dapat menangkap pola data untuk diprediksi. Setelah dilakukan seleksi fitur waktu yang dibutuhkan untuk prediksi menjadi kurang dari 1 detik saja karna beberapa fitur telah dihilangkan.

3.3. Evaluasi

Berikut perbandingan peningkatan dan penurunan akurasi, *precision*, *recall* dan f1-*score* pengujian data sebelum dilakukan seleksi fitur dengan pengujian data setelah dilakukan seleksi fitur. Lihat tabel 5 untuk evaluasi hasil.

Tabel 5. Evaluasi Hasil

Kernel	Akurasi		Presisi		Recall		F1-Score	
	Sebelum	Setelah	Sebelum	Setelah	Sebelum	Setelah	Sebelum	Setelah
Linear	98.95%	99.65%	97.71%	99.22%	100%	100%	98.84%	99.61%
Poly	94.74%	98.60%	93.80%	96.97%	94.53%	100%	94.16%	98.46%
RBF	93.68%	98.25%	92.97%	96.24%	92.97%	100%	92.97%	98.08%
Sigmoid	47.02%	12.98%	40.17%	0%	36.72%	0%	38.37%	0%

Berdasarkan pengujian data menggunakan empat kernel SVM, membuktikan bahwa metode seleksi fitur algoritma *firefly* menunjukan peningkatan nilai akurasi, *precision*, *recall* dan f1-*score* untuk ketiga

model kernel yang digunakan yaitu pada kernel *linear, polynomial* dan RBF, akan tetapi metode seleksi fitur algoritma *firefly* juga dapat menyebabkan penurunan akurasi, *precision*, *recall* dan f1-*score* untuk kernel sigmoid hal ini diakibatkan karena adanya *underfitting* karena setelah dilakukan seleksi fitur data menjadi terlalu sederhana sehingga kernel tidak dapat mempelajari dan menangkap pola data dengan benar. Selain itu kernel sigmoid sangat terpengaruh oleh pengurangan fitur karena kehilangan fitur dapat menghilangkan interaksi non-linear penting sehingga menyebabkan fungsi tanh menjadi tidak informatif dan model sulit membedakan kelas. Pemilihan fitur *gender* dan hemoglobin oleh algoritma *firefly* menunjukkan bahwa proses seleksi fitur berhasil mempertahankan fitur yang benar-benar relevan secara medis juga dapat menjadikan pola anemia yang lebih sederhana dan efisien terlebih untuk sistem prediksi dan klasifikasi yang ringan dan cepat. Penelitian ini menunjukkan bahwa kombinasi algoritma SVM dengan seleksi fitur menggunakan algoritma *firefly* mampu menghasilkan model klasifikasi dan prediksi yang efisien pada beberapa kernel dengan hanya menggunakan dua fitur terpilih namun memperoleh tingkat akurasi yang lebih tinggi. Hasil ini sejalan dengan studi terdahulu [34] yang menggunakan algoritma SVM dengan tambahan seleksi fitur dalam prediksi penyakit menghasilkan akurasi model yang tinggi. Dalam studi lain [35] dilakukan klasifikasi anemia dengan algoritma SVM tanpa seleksi fitur menghasilkan akurasi yang lebih rendah. Dengan demikian penelitian ini memiliki kontribusi dalam penggunaan seleksi fitur algoritma *firefly* untuk mengoptimalkan kinerja algoritma SVM pada data anemia sehingga menghasilkan solusi yang lebih ringan , cepat dan praktis untuk prediksi anemia berbasis *machine learning*.

4. KESIMPULAN

Setelah dilakukan seleksi fitur menggunakan algoritma *Firefly* didapatkan kombinasi fitur terbaik yaitu fitur ke 0 dan ke 1 dimana fitur tersebut yaitu “Gender” dan “Hemoglobin”. Akurasi prediksi setelah seleksi fitur pada *kernel linear* mengalami peningkatan sebesar 0.7% dari akurasi awal 98.95% menjadi 99.65%, *kernel polynomial* mengalami peningkatan sebesar 3.86% dari akurasi awal 94.74% menjadi 98.60%, *kernel* RBF mengalami peningkatan sebesar 4.57% dari akurasi awal 93.68% menjadi 98.25%, akan tetapi pada *kernel sigmoid* akurasi mengalami penurunan sebanyak 34.04% dari akurasi awal 47.02% menjadi 12.98%. Penggunaan FA efektif dalam mengidentifikasi fitur-fitur penting yang berdampak pada akurasi prediksi untuk tiga kernel SVM, sehingga dapat menghasilkan akurasi yang lebih tinggi. Seleksi fitur menjadi efektif jika menghasilkan kombinasi fitur yang tepat dan dapat menjadi tidak efektif jika menghasilkan kombinasi fitur yang tidak tepat. Saran yang dapat penulis berikan guna penelitian selanjutnya yaitu membandingkan kinerja FA dengan metode seleksi fitur lainnya seperti *Genetic Algorithm* (GA) atau *Particle Swarm Optimization* (PSO) untuk mengetahui efektivitas relatif dari *Firefly Algorithm* (FA). Menguji model pada berbagai algoritma klasifikasi selain SVM, seperti *Random Forest* atau *Neural Networks*, untuk mengetahui kemungkinan peningkatan kinerja prediksi.

DAFTAR PUSTAKA

[1] D. E. Yanti, L. Framesti, and A. Desiani, “Perbandingan Algoritma C4.5 Dan Svm Dalam Klasifikasi Penyakit Anemia,” *J. Inform. Polinema*, vol. 9, no. 4, pp. 427–434, 2023, doi: 10.33795/jip.v9i4.1381.

[2] T. Meischl *et al.*, “Anaemia is independently associated with mortality in patients with hepatocellular carcinoma,” *ESMO Open*, vol. 9, no. 6, p. 103593, 2024, doi: 10.1016/j.esmoop.2024.103593.

[3] D. Chandra, V. Capoor, A. Maitri, and R. Naithani, “Autoimmune Hemolytic Anemia in Children,” *Pediatr. Hematol. Oncol. J.*, 2024, doi: 10.1016/j.phoj.2024.08.002.

[4] S. R. Nadhiroh, A. R. Hasugian, Nurhayati, A. D. Muthiah, and A. N. P. A. Putri, “Model development for anemia prediction in pregnancy,” *Clin. Epidemiol. Glob. Heal.*, vol. 28, no. June, p. 101654, 2024, doi: 10.1016/j.cegh.2024.101654.

[5] M. Yackobovitch-Gavan, D. Ben-Hefer, I. Feldhamer, and J. Meyerovitch, “The association between infantile microcytic anemia and attention deficit hyperactivity disorder, a case-control study,” *Heliyon*, vol. 10, no. 12, p. e33430, 2024, doi: 10.1016/j.heliyon.2024.e33430.

[6] J. Pacheco-Aranibar, K. Diaz-Rodriguez, R. Zapana-Begazo, S. Criollo-Arteaga, J. A.

Villanueva-Salas, and J. C. Bernabe-Ortiz, "Intestinal microbiota dataset revealed by high-throughput sequencing of 16S rRNA in children with anemia in southern Peru," *Data Br.*, vol. 55, p. 110681, 2024, doi: 10.1016/j.dib.2024.110681.

[7] E. Tasia, R. Zaid, I. Zarier, S. Kenia, and P. Loka, "Klasifikasi Penyakit Gagal Jantung Menggunakan Supervised Learning," *Sentimas*, pp. 1–7, 2023, [Online]. Available: https://journal.irpi.or.id/index.php/sentimas/article/view/535

[8] M. Azhari, Z. Situmorang, and R. Rosnelly, "Perbandingan Akurasi, Recall, dan Presisi Klasifikasi pada Algoritma C4.5, Random Forest, SVM dan Naive Bayes," *J. Media Inform. Budidarma*, vol. 5, no. 2, p. 640, 2021, doi: 10.30865/mib.v5i2.2937.

[9] R. G. Wardhana, G. Wang, and F. Sibuea, "Penerapan Machine Learning Dalam Prediksi Tingkat Kasus Penyakit Di Indonesia," *J. Inf. Syst. Manag.*, vol. 5, no. 1, pp. 40–45, 2023, doi: 10.24076/joism.2023v5i1.1136.

[10] S. Samantaray, A. Sahoo, and F. Baliarsingh, "Groundwater level prediction using an improved SVR model integrated with hybrid particle swarm optimization and firefly algorithm," *Clean. Water*, vol. 1, no. December 2023, p. 100003, 2024, doi: 10.1016/j.clwat.2024.100003.

[11] E. Pusporani, S. Qomariyah, and I. Irhamah, "Klasifikasi Pasien Penderita Penyakit Liver dengan Pendekatan Machine Learning," *Inferensi*, vol. 2, no. 1, p. 25, 2019, doi: 10.12962/j27213862.v2i1.6810.

[12] M. S. Reza, U. Hafsha, R. Amin, R. Yasmin, and S. Ruhi, "Improving SVM performance for type II diabetes prediction with an improved non-linear kernel: Insights from the PIMA dataset," *Comput. Methods Programs Biomed. Updat.*, vol. 4, no. August, p. 100118, 2023, doi: 10.1016/j.cmpbup.2023.100118.

[13] M. Zhang, M. Treder, D. Marshall, and Y. Li, "Explaining the predictions of kernel SVM models for neuroimaging data analysis," *Expert Syst. Appl.*, vol. 251, no. March, p. 123993, 2024, doi: 10.1016/j.eswa.2024.123993.

[14] I. Anshory *et al.*, "Optimization DC-DC boost converter of BLDC motor drive by solar panel using PID and firefly algorithm," *Results Eng.*, vol. 21, no. March 2023, p. 101727, 2024, doi: 10.1016/j.rineng.2023.101727.

[15] Y. A. Sari, A. G. Hapsani, S. Adinugroho, L. Hakim, and S. Mutrofin, "Preprocessing of Skin Images and Feature Selection for Early Stage of Melanoma Detection using Color Feature Extraction," *Int. J. Artif. Intell. Res.*, vol. 4, no. 2, p. 95, 2021, doi: 10.29099/ijair.v4i2.165.

[16] S. Jalota and M. Suthar, "Modelling of Marshall stability of polypropylene fibre reinforced asphalt concrete using support vector machine and artificial neural network," *International Journal of Transportation Science and Technology*, Tongji University and Tongji University Press, 2024. doi: 10.1016/j.ijtst.2024.08.001.

[17] S. Dwiasnati and Y. Devianto, "Optimasi Prediksi Bencana Banjir menggunakan Algoritma SVM untuk penentuan Daerah Rawan Bencana Banjir," *Pros. SISFOTEK*, pp. 202–207, 2021, [Online]. Available: http://seminar.iaii.or.id/index.php/SISFOTEK/article/view/283

[18] N. Maulidah, R. Supriyadi, D. Y. Utami, F. N. Hasan, A. Fauzi, and A. Christian, "Prediksi Penyakit Diabetes Melitus Menggunakan Metode Support Vector Machine dan Naive Bayes," *Indones. J. Softw. Eng.*, vol. 7, no. 1, pp. 63–68, 2021, doi: 10.31294/ijse.v7i1.10279.

[19] H. S. W. Hovi, A. Id Hadiana, and F. Rakhmat Umbara, "Prediksi Penyakit Diabetes Menggunakan Algoritma Support Vector Machine (SVM)," *Informatics Digit. Expert*, vol. 4, no. 1, pp. 40–45, 2022, doi: 10.36423/index.v4i1.895.

[20] Y. Simamora, "Optimasi Rekonfigurasi Jaringan Distribusi Tegangan Menengah Kota Medan Menggunakan Algoritma Kunang – Kunang (Firefly Algorithm)," *Sutet*, vol. 13, no. 2, pp. 82–90, 2024, doi: 10.33322/sutet.v13i2.2240.

[21] I. Septiyafi, H. Suprajitno, and A. B. Pratiwi, "Penerapan Algoritma Kunang-Kunang pada Open Vehicle Routing Problem (OVRP)," *Contemp. Math. Appl.*, vol. 1, no. 1, pp. 46–55, 2019, doi: 10.20473/conmatha.v1i1.14774.

[22] I. Czarnowski, "Firefly algorithm for instance selection," *Procedia Comput. Sci.*, vol. 192, pp.

2269–2278, 2021, doi: 10.1016/j.procs.2021.08.240.

[23] R. Pujianto, D. Yusup, and T. N. Padilah, "Analisis Sentimen Opini Publik Tentang Vaksin Booster Menggunakan Metode Support Vector Machine dan firefly Algorithm," *J. Ilm. Wahana Pendidikan, Desember*, vol. 2022, no. 23, pp. 363–373, 2022, [Online]. Available: https://doi.org/10.5281/zenodo.7397891

[24] Styawati, Andi Nurkholis, Zaenal Abidin, and Heni Sulistiani, "Optimasi Parameter Support Vector Machine Berbasis Algoritma Firefly Pada Data Opini Film," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 5, pp. 904–910, 2021, doi: 10.29207/resti.v5i5.3380.

[25] M. Lutfi, H. S. Rizal, M. Hasyim, M. F. Amrulloh, and Z. N. Saadah, "Feature Extraction and Naïve Bayes Algorithm for Defect Classification of Manalagi Apples," *J. Phys. Conf. Ser.*, vol. 2394, no. 1, 2022, doi: 10.1088/1742-6596/2394/1/012014.

[26] N. Bacanin *et al.*, "Quasi-reflection learning arithmetic optimization algorithm firefly search for feature selection," *Heliyon*, vol. 9, no. 4, pp. 1–11, 2023, doi: 10.1016/j.heliyon.2023.e15378.

[27] N. Bacanin, K. Venkatachalam, T. Bezdan, M. Zivkovic, and M. Abouhawwash, "A novel firefly algorithm approach for efficient feature selection with COVID-19 dataset," *Microprocess. Microsyst.*, vol. 98, no. January, 2023, doi: 10.1016/j.micpro.2023.104778.

[28] A. P. Masoumi, A. R. Tavakolpour-Saleh, and V. Bagherian, "Performance investigation of an active free-piston Stirling engine using artificial neural network and firefly optimization algorithm," *Heliyon*, vol. 10, no. 7, pp. 1–19, 2024, doi: 10.1016/j.heliyon.2024.e28387.

[29] T. Yuan, Y. Mu, T. Wang, Z. Liu, and A. Pirouzi, "Using firefly algorithm to optimally size a hybrid renewable energy system constrained by battery degradation and considering uncertainties of power sources and loads," *Heliyon*, vol. 10, no. 7, p. e26961, 2024, doi: 10.1016/j.heliyon.2024.e26961.

[30] S. Larabi Marie-Sainte and N. Alalyani, "Firefly Algorithm based Feature Selection for Arabic Text Classification," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 32, no. 3, pp. 320–328, 2020, doi: 10.1016/j.jksuci.2018.06.004.

[31] S. Rabbani, D. Safitri, N. Rahmadhani, A. A. F. Sani, and M. K. Anam, "Perbandingan Evaluasi Kernel SVM untuk Klasifikasi Sentimen dalam Analisis Kenaikan Harga BBM," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 3, no. 2, pp. 153–160, Oct. 2023, doi: 10.57152/malcom.v3i2.897.

[32] X. Wang, Z. Yan, Y. Zeng, X. Liu, X. Peng, and H. Yuan, "Research on correlation factor analysis and prediction method of overhead transmission line defect state based on association rule mining and RBF-SVM," *Energy Reports*, vol. 7, pp. 359–368, 2021, doi: 10.1016/j.egyr.2021.01.058.

[33] M. Han, S. Soradi-Zeid, T. Anwlnkom, and Y. Yang, "Firefly algorithm-based LSTM model for Guzheng tunes switching with big data analysis," *Heliyon*, vol. 10, no. 12, p. e32092, 2024, doi: 10.1016/j.heliyon.2024.e32092.

[34] W. S. Dharmawan, "KOMPARASI ALGORITMA KLASIFIKASI SVM-PSO DAN C4.5-PSO DALAM PREDIKSI PENYAKIT JANTUNG," *J. Inform. Manaj. dan Komput.*, vol. 13, no. 2, pp. 31–41, 2021.

[35] D. E. Yanti, L. Framesti, and A. Desiani, "PERBANDINGAN ALGORITMA C4.5 DAN SVM DALAM KLASIFIKASI PENYAKIT ANEMIA," *J. Inform. Polinema*, vol. 9, no. 4, pp. 427–434, Aug. 2023, doi: 10.33795/jip.v9i4.1381.