

**Prediksi Aktivitas Inhibitor VEGFR-2 sebagai Agen
Anti-Kanker Menggunakan Support Vector Machine dan
Grey Wolf Optimizer**

Proposal Tugas Akhir

Kelas MK Penulisan Proposal (CCH4A3)

1301223212

Andira Lupita



Program Studi Sarjana S1 Informatika

Fakultas Informatika

Universitas Telkom

Bandung

2024

Lembar Persetujuan

Prediksi Aktivitas Inhibitor VEGFR-2 sebagai Agen Anti-Kanker Menggunakan Support Vector Machine dan Grey Wolf Optimizer

Prediction of VEGFR-2 Inhibitor Activity as an Anti-Cancer Agent Using Support Vector Machine and Grey Wolf Optimizer

NIM :1301223212

Andira Lupita

Proposal ini diajukan sebagai usulan pembuatan tugas akhir pada
Program Studi Sarjana S1 Informatika
Fakultas Informatika Universitas Telkom

Bandung, 19 Mei 2025

Menyetujui

Isman Kurniawan, S.Pd., M.Si., M.Sc., Ph.D.

NIP: 15870066

ABSTRAK

Kanker merupakan penyebab utama kematian di dunia dengan estimasi 1,95 juta kasus baru pada tahun 2023 di Amerika Serikat, di mana angiogenesis yang dimediasi oleh VEGFR-2 berperan penting dalam pertumbuhan dan penyebaran tumor. Salah satu strategi terapi adalah penggunaan inhibitor VEGFR-2 untuk menghambat pembentukan pembuluh darah baru, namun efektivitasnya masih menghadapi tantangan seperti resistensi obat dan efek samping. Oleh karena itu, penelitian ini bertujuan mengembangkan model prediksi aktivitas inhibitor VEGFR-2 menggunakan pendekatan *in silico* berbasis machine learning dengan mengombinasikan Support Vector Machine (SVM) dan algoritma optimasi Grey Wolf Optimizer (GWO). SVM dipilih karena kemampuannya dalam klasifikasi berbasis pola kimia kompleks, sementara GWO digunakan untuk melakukan feature selection dan tuning hyperparameter guna meningkatkan akurasi prediksi. Dataset yang digunakan terdiri dari 118 senyawa kandidat inhibitor VEGFR-2 dalam format SMILES yang diperoleh dari Taylor & Francis. Tahapan penelitian meliputi *pre-processing* data, seleksi fitur menggunakan GWO, pembangunan model prediktif dengan SVM, serta evaluasi performa model menggunakan metrik akurasi, R^2 , dan MAPE. Berdasarkan studi terdahulu, kombinasi SVM dan GWO telah terbukti mampu meningkatkan prediktabilitas dalam konteks klasifikasi senyawa bioaktif. Hipotesis awal dari penelitian ini adalah bahwa integrasi GWO dalam proses seleksi fitur dan tuning hyperparameter akan meningkatkan akurasi model SVM secara signifikan dalam memprediksi aktivitas inhibitor VEGFR-2. Dengan demikian, penelitian ini diharapkan dapat berkontribusi dalam mempercepat proses penemuan kandidat obat anti-kanker yang lebih efektif dan aman.

Kata kunci: VEGFR-2, inhibitor, kanker, SVM, GWO, *in silico*.

1. PENDAHULUAN

1.1. Latar Belakang

Kanker merupakan salah satu penyebab kematian utama di dunia. Pada tahun 2023 diperkirakan akan ada sekitar 1.95 juta kasus kanker baru, atau setara dengan sekitar 5.370 kasus setiap harinya di Amerika Serikat[1]. Penyakit ini disebabkan oleh perubahan genetik yang menyebabkan pertumbuhan sel abnormal (tumor) secara tidak terkendali dan dapat menyebar ke organ lain (metastasis)[2][3]. Salah satu faktor yang menyebabkan pertumbuhan tumor secara masif adalah karena adanya pembentukan pembuluh darah baru (angiogenesis) melalui aktivitas *vascular endothelial growth factors* (VEGF). Reseptor utama dari VEGF adalah *vascular endothelial growth factors receptor-2* (VEGFR-2) yang memainkan peran utama dalam angiogenesis[4]. VEGFR-2 sebagai reseptor yang memberikan sel darah baru nutrisi dan oksigen, tetapi terlalu banyak VEGFR-2 membantu menyebarkan metastasis tumor[5]. Oleh karena itu, penghambatan VEGFR-2 melalui penggunaan inhibitor VEGFR-2 sebagai anti kanker menjadi salah satu strategi dalam pengobatan kanker[6]. Dengan menghalangi sinyal VEGF yang berikatan dengan VEGFR-2, maka inhibitor dapat mengganggu aliran oksigen dan nutrisi ke tumor, sehingga menghambat perkembangannya. Beberapa inhibitor VEGFR-2, seperti ramucirumab telah disetujui untuk terapi kanker dan menunjukkan efektivitas dalam menghambat angiogenesis tumor[7]. Namun, tantangan seperti resistensi obat dan efek samping kardiovaskular mendorong penelitian lebih lanjut untuk mengembangkan inhibitor VEGFR-2 yang lebih aman dan efektif[8].

Untuk mempercepat penemuan produk obat, pendekatan *in silico* berbasis *machine learning* semakin banyak digunakan dalam proses *drug discovery*. Ding et al. pada tahun 2024 menggabungkan pendekatan machine learning seperti Topomer CoMFA dan algoritma K-Nearest

Neighbors (KNN) serta analisis molecular docking untuk memprediksi inhibitor VEGFR-2, dan berhasil mengidentifikasi lima senyawa potensial dengan akurasi prediksi di atas 80%[9]. Dalam konteks prediksi obat, pada tahun 2021 Masoomi Sefiddashti et al. telah membandingkan model Multiple Linear Regression (MLR) dan Artificial Neural Network (ANN) untuk memprediksi aktivitas 33 turunan fuoropyrimidine dan tienopyrimidine terhadap VEGFR-2, dengan hasil bahwa ANN menunjukkan nilai R^2 sebesar 0,998, jauh lebih tinggi dibandingkan MLR[10]. Pada penelitian yang dilakukan V. Napolitano et al. di tahun 2022 dengan melakukan pengembangan model QSAR studi ini menunjukkan nilai R^2 sebesar 0,829 yang memperlihatkan validitas prediktif yang baik terhadap potensi senyawa sebagai inhibitor VEGFR-2[11]. Pada tahun 2022, Salimi et al. melakukan penelitian menggunakan model Random Forest untuk melakukan klasifikasi dan prediksi aktivitas inhibitor VEGFR dengan mengidentifikasi nilai pIC_{50} menghasilkan nilai sebesar 7,23 dan 7,31 yang sebanding dengan aktivitas inhibitor VEGFR yang telah disetujui seperti pazopanib dan sorafenib[12]. Selain itu terdapat juga penelitian pada tahun 2023 oleh Arabi et al yang menunjukkan bahwa SVM dengan kernel radial basis function (RBF) mampu mengklasifikasikan senyawa sebagai inhibitor atau non-inhibitor VEGFR-2 dengan akurasi 82,4%, menggunakan data dari basis data BindingDB[13]. SVM dapat mengklasifikasikan koneksi berdasarkan pola kimia yang kompleks, namun kinerjanya sangat tergantung pada parameter yang dipilih[2]. Berdasarkan hal tersebut dibutuhkan metode optimasi untuk mengoptimalkan parameter ini. Salah satu algoritma optimasi yang efektif adalah dengan menggunakan algoritma metaheuristik yaitu Grey Wolf Optimizer (GWO). Algoritma GWO digunakan untuk memecahkan masalah optimasi dengan jumlah variabel yang besar dan kendala yang kompleks. Berdasarkan penelitian Chen et al. (2023) dalam konteks optimasi model prediktif, algoritma metaheuristik seperti Multi-Objective Grey Wolf Optimizer (MOGWO) menawarkan keunggulan dalam pemilihan parameter optimal[14]. Kombinasi SVM dan

GWO telah terbukti meningkatkan prediktabilitas dalam berbagai studi, termasuk klasifikasi dan regresi data kompleks.

Tujuan dari penelitian ini adalah untuk mengembangkan model prediksi aktivitas inhibitor VEGFR-2 sebagai agen anti-kanker menggunakan pendekatan *in silico* menggunakan kombinasi SVM dan GWO. Diharapkan integrasi kedua metode ini dapat diperoleh model prediktif yang lebih akurat dan efisien dalam mengklasifikasikan senyawa sebagai inhibitor VEGFR-2, dengan demikian dapat mempercepat proses penemuan agen anti-kanker. Diharapkan penelitian ini dapat menjawab tantangan terhadap keterbatasan inhibitor yang ada seperti resistensi dan efek samping serta untuk mendukung pengembangan terapi kanker yang lebih efektif dan aman.

1.2. Perumusan Masalah

- 1) Bagaimana efektivitas algoritma GWO dalam melakukan *feature selection* terhadap dataset molekuler guna meningkatkan akurasi prediksi aktivitas inhibitor VEGFR-2?
- 2) Bagaimana pengaruh hyperparameter tuning terhadap performa model SVM dalam memprediksi aktivitas inhibitor VEGFR-2?
- 3) Bagaimana evaluasi performa model SVM terhadap aktivitas inhibitor VEGFR-2 setelah diterapkannya seleksi fitur dan optimasi hyperparameter dengan GWO?

1.3. Tujuan

- 1) Menganalisis efektivitas *feature selection* menggunakan GWO dalam melakukan prediksi aktivitas inhibitor VEGFR-2 sebagai agen anti-kanker.
- 2) Menganalisis pengaruh hyperparameter tuning terhadap performa SVM dalam memprediksi aktivitas inhibitor VEGFR-2.
- 3) Mengevaluasi performa model SVM dalam memprediksi aktivitas inhibitor VEGFR-2 setelah diterapkannya seleksi fitur dan optimasi

hyperparameter menggunakan GWO.

1.4. Rencana Kegiatan

1) *Literature Review*

Tahap ini akan dilakukan review terkait topik penelitian ini. Studi literatur akan menjadi landasan serta referensi dalam penulisan penelitian ini terkait metode yang akan digunakan dalam membangun model.

2) Pencarian *Dataset*

Pencarian dataset yang akan digunakan dalam penelitian ini. Dataset yang terpilih akan menjadi data yang digunakan dalam penelitian ini.

3) *Pre-Processing Data*

Langkah selanjutnya yang dilakukan adalah *pre-processing* data. Data akan diidentifikasi terlebih dahulu kemudian akan didapatkan data yang bersih dan baik untuk digunakan. Kemudian data yang telah bersih akan dapat diproses untuk tahap selanjutnya.

4) Pembuatan Model

Dataset yang sudah siap digunakan akan masuk ke dalam tahap pembuatan model. Pada tahap ini, data akan dibagi menjadi data training dan testing. Setelahnya akan diberikan beberapa *training techiques* sebelum akhirnya dievaluasi.

5) Evaluasi Model

Setelah mendapatkan model untuk mengetahui seberapa baik model yang telah dibuat akan dilakukan evaluasi, Pada tahap ini, model akan dievaluasi menggunakan beberapa metrik evaluasi.

6) Menulis Laporan

Tahap terakhir yaitu penulisan laporan yang merupakan dokumentasi dalam bentuk tulisan dari setiap kegiatan yang dilakukan dalam penelitian.

1.5. Jadwal Kegiatan

| Kegiatan | Bulan | | | | | |
|----------------------------|-------|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| <i>Literature Review</i> | | | | | | |
| Pencarian <i>Dataset</i> | | | | | | |
| <i>Pre-processing</i> Data | | | | | | |
| Pembuatan Model | | | | | | |
| Evaluasi Hasil | | | | | | |
| Penulisan Laporan | | | | | | |

2. KAJIAN PUSTAKA

2.1. Studi Terkait

Penelitian mengenai prediksi aktivitas inhibitor VEGFR-2 telah banyak dilakukan dengan pendekatan *in silico* seperti halnya berbagai metode pembelajaran mesin. Studi-studi ini merupakan sebuah referensi yang di gunakan dalam penelitian.

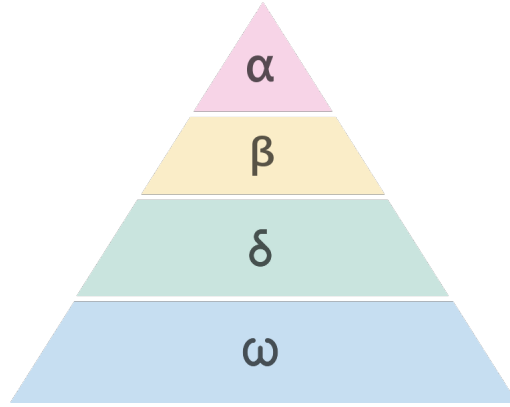
Penelitian prediksi aktivitas inhibitor VEGFR-2 telah banyak dilakukan menggunakan pendekatan *in silico* berbasis QSAR, *machine learning*, dan molecular docking. Masoomi et al. pada tahun 2021 membandingkan performa MLR dan ANN pada 33 turunan furopyrimidine dan tienopyrimidine, menunjukkan bahwa ANN lebih unggul dalam menangkap hubungan nonlinier dengan nilai R^2 sebesar 0,998[10]. Studi serupa oleh V. Napolitano et al. di tahun 2022 menggunakan QSAR untuk aktivitas inhibitor VEGFR-2 yang valid secara statistik nilai R^2 sebesar 0,829 menegaskan potensi prediktif pendekatan terintegrasi dalam pemodelan prediktif senyawa antikanker[11]. Sementara itu, Salimi et al. pada tahun 2022 yang menggunakan model Random Forest berhasil memprediksi aktivitas inhibitor pada senyawa baru dengan nilai sebesar 7,23 dan 7,31 yang setara dengan inhibitor yang sudah ada[12]. Selain pendekatan klasik, metode *machine learning* seperti SVM mulai menunjukkan potensi tinggi dalam klasifikasi senyawa bioaktif. Arabi et al. pada tahun 2023 menggunakan SVM dengan kernel RBF dan data dari BindingDB untuk mengklasifikasikan senyawa sebagai inhibitor VEGFR-2 dengan akurasi sebesar 82,4%[13]. Di sisi lain, Ding et al. pada 2024 menggabungkan Topomer CoMFA, KNN, dan molecular docking untuk mengidentifikasi lima senyawa kandidat dengan akurasi prediksi yang konsisten di atas 80%[9]. Pendekatan multi-model ini menunjukkan bahwa kombinasi teknik dapat meningkatkan akurasi dan keandalan dalam prediksi aktivitas biologis senyawa. Namun, efektivitas SVM sangat bergantung pada pemilihan parameter optimal. Untuk itu, algoritma optimasi seperti GWO digunakan untuk menyempurnakan performa model. Penelitian yang dilakukan Chen et al. pada tahun 2023 menunjukkan bahwa GWO mampu mengatasi masalah optimasi multi-

objektif dalam pengobatan kanker[14].

Seiring perkembangan teknologi komputasi, pendekatan *in silico* mulai diadopsi untuk mempercepat dan menyederhanakan proses penemuan obat. Pendekatan *in silico* dalam studi oleh Yousef et al. pada tahun 2022 terbukti efektif dalam menyaring dan memvalidasi kandidat senyawa sebelum dilanjutkan ke uji biologis, sehingga mempercepat proses penemuan obat dan mengurangi ketergantungan pada metode *in vitro* yang lebih mahal dan memakan waktu[15]. Dalam penelitian ini, pendekatan *in silico* dikombinasikan dengan algoritma pembelajaran mesin Support Vector Machine (SVM) dan Grey Wolf Optimizer (GWO) untuk memprediksi aktivitas inhibitor berdasarkan data molekuler. Penggabungan SVM dan GWO ini menjadi pendekatan yang menjanjikan dalam mengembangkan model prediktif aktivitas inhibitor VEGFR-2 yang lebih efisien dan andal.

2.2. Grey Wolf Algorithm

Grey Wolf Optimizer (GWO) merupakan algoritma meta-heuristik yang meniru perilaku sosial dan kebiasaan berburu serigala abu-abu (*grey wolf*) di alam liar. Algoritma meniru struktur hierarki sosial yang terdiri dari empat peran utama: alfa (α) sebagai pengambil keputusan, beta (β) sebagai fasilitator, delta (δ) sebagai pelaksana, dan omega (ω) sebagai anggota dengan peran terendah yakni sebagai penjaga harmoni kelompok[16]. Penelitian yang dilakukan oleh Liu et al. dalam algoritma ini, peran-peran tersebut digunakan untuk merepresentasikan solusi dalam proses optimasi, di mana solusi α dianggap paling optimal, β sebagai solusi terbaik kedua, dan seterusnya hingga ω sebagai solusi terakhir yang kurang optimal[17].



Gambar 1. Ilustrasi Hirarki *Grey Wolf*

Dalam penerapannya, GWO mengadopsi tiga tahapan utama dalam proses pencarian solusi optimal, yakni: (1) mendekati dan mengelilingi target, (2) perilaku berburu, serta (3) menyerang dan mengeksploitasi target. Strategi ini mencerminkan bagaimana serigala abu-abu secara kolektif mengepung dan memburu mangsanya dengan mengandalkan struktur sosial mereka. Pendekatan ini memungkinkan pengumpulan solusi yang dipilih secara acak berdasarkan preferensi hierarki sosial dalam algoritma, menjadikannya sangat efektif dalam proses pencarian solusi global pada permasalahan optimasi.

Tahapan tersebut dimodelkan secara sistematis menggunakan pendekatan vektor. Proses menyerang mangsa dimulai dengan menghitung jarak posisi antara serigala dan mangsa menggunakan:

$$\vec{D} = |\vec{C} \cdot \vec{X}_{victim}(t) - \vec{X}_{wolf}(t)| \quad (1)$$

$$\vec{X}_{wolf}(t+1) = \vec{X}_{victim}(t) - \vec{A} \cdot \vec{D} \quad (2)$$

Dimana \vec{A} dan \vec{C} adalah vektor koefisien acak yang mengontrol intensitas eksplorasi dan eksploitasi, sedangkan $\vec{X}_{wolf}(t)$ adalah posisi serigala dan $\vec{X}_{victim}(t)$ adalah posisi mangsa. Nilai $|\vec{A}|$ menjadi penentu arah gerakan serigala, sehingga apabila $|\vec{A}| > 1$, maka serigala menjauh dari mangsa (eksplorasi), sementara apabila $|\vec{A}| = 1$ adalah sebaliknya yaitu mendekat ke mangsa (eksploitasi) yang mana hal ini dianggap optimal. Perilaku berburu dimodelkan dengan memperbaharui posisi serigala berdasarkan posisi tiga pemimpin utama

(α, β, δ) dengan rumus:

$$\vec{D}_\alpha = |\vec{C}_1 \cdot \vec{X}_\alpha - \vec{X}_{wolf}|, \quad \vec{X}_1 = \vec{X}_\alpha - \vec{A}_1 \cdot \vec{D}_\alpha \quad (3)$$

$$\vec{D}_\beta = |\vec{C}_2 \cdot \vec{X}_\beta - \vec{X}_{wolf}|, \quad \vec{X}_2 = \vec{X}_\beta - \vec{A}_2 \cdot \vec{D}_\beta \quad (4)$$

$$\vec{D}_\delta = |\vec{C}_3 \cdot \vec{X}_\delta - \vec{X}_{wolf}|, \quad \vec{X}_3 = \vec{X}_\delta - \vec{A}_3 \cdot \vec{D}_\delta \quad (5)$$

$$\vec{X}_{wolf}(t+1) = \frac{\vec{X}_1 + \vec{X}_2 + \vec{X}_3}{3} \quad (6)$$

Berikut merupakan algoritma GWO[[16]]:

Definisi: N_p : ukuran populasi; T : jumlah maksimum iterasi

inisialisasi populasi agen pencarian $X_i (i = 1, 2, \dots, N_p)$;

inisialisasi hyper-paramater a, A, dan C;

while $t < T$ do

pemeriksaan batas untuk setiap agen pencarian;

hitung nilai fitness dari $X_i (i = 1, 2, \dots, N_p)$;

identifikasi X_α, X_β , dan X_δ ;

for i dari 1 sampai N_p do

perbaharui a, A, dan C;

perbaharui agen pencarian saat ini (X_i) menggunakan persamaan

end

$t = t + 1$

end

kembalikan X_α

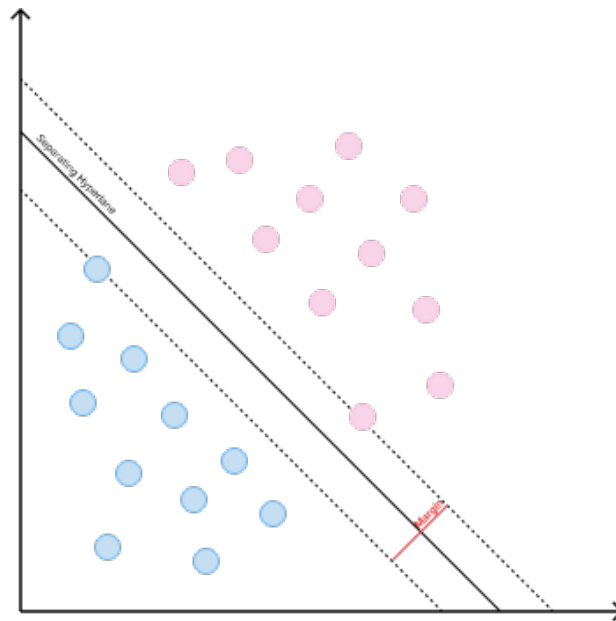
Algoritma 1. Pseudocode GWO

2.3. Support Vector Machine

Support Vector Machine (SVM) adalah metode pembelajaran statistik yang digunakan untuk klasifikasi dan regresi, termasuk daerah yang memprediksi

aktivitas biologis koneksi[2]. Dalam arsitektur SVM terdapat *hyperplane* yang memisahkan dua kelas data dengan margin maksimum yang mengarah ke model dengan fungsi generalisasi yang baik[2]. Ilustrasi dari *hyperplane* pada SVM terdapat pada Gambar 2. *Hyperlane* tersebut dibangun berdasarkan persamaan:

$$w \cdot x + b = 0 \quad (7)$$



Gambar 2. Ilustrasi dari SVM *Hyperplane*

Tujuan utama SVM adalah tidak hanya menemukan pemisah kelas yang tepat, tetapi juga memaksimalkan jarak margin antara *hyperplane* dan data tiap kelas. Semakin besar margin, semakin baik kemampuan model melakukan generalisasi terhadap data baru. Dalam struktur SVM, *support vectors* memainkan peran kunci. *Support vectors* adalah titik-titik data yang berada paling dekat dengan *hyperplane* dan menentukan posisi *hyperplane* tersebut. Margin yang diukur dari jarak antara *support vectors* dan *hyperplane* harus diperbesar untuk mengoptimalkan performansi dari model.

Fungsi kernel memungkinkan SVM untuk memetakan data dari ruang input ke ruang dimensi yang lebih tinggi, secara efektif memisahkan data yang tidak dapat dipisahkan secara linear. Pemilihan kernel yang sesuai sangat berpengaruh terhadap performa SVM, terutama pada kasus-kasus kompleks seperti deteksi kanker atau klasifikasi citra medis. Beberapa jenis kernel yang biasanya digunakan

dalam SVM adalah Polynomial, Radial Basis Function, Mahalanobis, dan Sigmoid[2]. Berikut beberapa persamaan yang dapat digunakan:

1) Polynomial

$$K(x_i, x_j) = (\gamma x_i^T x_j + 1)^d, \quad \gamma > 0 \quad (8)$$

2) Radial Basis Function (RBF)

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2) \quad (9)$$

3) Linear

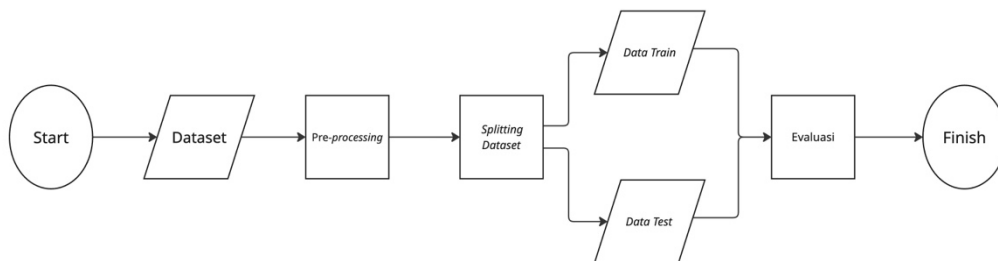
$$K(x_i, x_j) = x_i^T x_j \quad (10)$$

Pemilihan kernel yang sesuai sangat berpengaruh terhadap performa SVM, terutama pada kasus-kasus kompleks seperti deteksi kanker atau klasifikasi citra medis. Penyesuaian parameter yang cermat dapat meningkatkan performa model SVM secara signifikan dan menghasilkan prediksi yang lebih akurat dan andal[2].

3. PERANCANGAN SISTEM

Sistem yang Dibangun

Penelitian ini terdiri dari beberapa tahapan seperti yang terlihat pada Gambar (3). Proses dimulai dengan tahap input dataset, kemudian dilanjutkan ke tahapan berikutnya sesuai dengan alur yang ditampilkan pada Gambar (3). Proses pembangunan sistem diawali dengan pengumpulan dataset yang kemudian melalui tahap *pre-processing* untuk membersihkan dan menyiapkan data agar layak digunakan. Setelah itu, dataset dibagi menjadi dua bagian yaitu *data train* dan *data test*. *Data train* digunakan untuk melatih model, sedangkan *data test* akan digunakan untuk menguji performa model yang telah dibangun. Sebelum model dilatih, dilakukan seleksi fitur untuk memilih atribut yang paling relevan guna meningkatkan efisiensi dan akurasi sistem. Selanjutnya, model dibangun menggunakan data train dan fitur terpilih, lalu dievaluasi menggunakan data test dengan mengukur metrik performa seperti akurasi atau F1-score. Setelah evaluasi selesai, proses sistem pun berakhir.



Gambar 3. Flowchart Penelitian

3.1. Dataset

Pada penelitian ini, dataset yang digunakan diambil dari situs Taylor and Francis dari penelitian “In silico insights into design of novel VEGFR-2 inhibitors: SMILES-based QSAR modelling, and docking studies on substituted benzo-fused heteronuclear derivatives”. Dataset yang digunakan terdiri dari 118 senyawa kandidat inhibitor VEGFR-2 yang diperoleh dalam format SMILES, dengan nilai aktivitas biologis pIC50 yang berkisar antara 5.022 hingga 8.959. Berdasarkan nilai pIC50, senyawa-senyawa ini dikelompokkan ke dalam empat kategori berdasarkan potensi inhibitorynya. Terdapat 22 senyawa dengan pIC50 > 8 yang menunjukkan

potensi inhibitor sangat tinggi, 26 senyawa dengan pIC50 antara 7 dan 8 yang memiliki potensi inhibitor tinggi, 65 senyawa dengan pIC50 antara 6 dan 7 yang termasuk dalam kategori potensi inhibitor sedang, dan 5 senyawa dengan pIC50 < 6 yang termasuk dalam kategori potensi inhibitor rendah. Struktur kimia senyawa-senyawa ini meliputi berbagai turunan heterosiklik seperti benzimidazol, oksadiazol, dan tiazol yang memiliki berbagai substituen seperti halogen, gugus trifluorometil, dan gugus karbonil. Untuk tujuan pengembangan model prediksi, dataset ini dibagi menggunakan tiga skema pemisahan (Split1, Split2, Split3), dengan simbol "+" untuk set positif, "-" untuk set negatif, "*" untuk *data test*, "#" untuk set validasi, dan kosong untuk senyawa yang tidak ditugaskan. Rasio pemisahan data adalah 80% untuk *data train* dan 20% untuk *data test* yang memungkinkan penerapan validasi silang untuk mengoptimalkan performa model prediktif dalam penelitian ini.

3.2. Seleksi Fitur

Pada penelitian ini, algoritma GWO digunakan sebagai metode optimasi metaheuristik yang dapat melakukan pencarian global secara efektif dalam ruang fitur yang besar dan kompleks. Penelitian ini diimplementasikan menggunakan pustaka Niapy untuk mengimplementasikan GWO yang dapat meningkatkan performa regresi model SVM. Tujuan dari seleksi fitur adalah untuk mencari subset fitur yang optimal dalam meningkatkan performa model dan mengurangi kompleksitas model dengan membatasi jumlah fitur. Pada Tabel 1 terdapat parameter utama yang digunakan dalam implementasi ini adalah ukuran populasi (*population_size*) dan jumlah iterasi (*max_iter*).

Tabel 1. Parameter Optimasi GWO

| Parameter | Value | Deskripsi |
|-----------------|-----------------|-----------------------------------|
| Population_size | [5, 10, 15, 20] | Ukuran populasi agen (serigala) |
| Max_iter | 10 | Jumlah maksimum iterasi algoritma |

Setiap solusi direpresentasikan sebagai vektor angka real (kontinu) berdimensi sama dengan jumlah fitur awal yang kemudian nilai real tersebut diubah menjadi 0 atau 1 (binarisasi) menggunakan ambang batas tertentu. Subset fitur yang dipilih akan dievaluasi dengan melatih model dari SVM regresi dan untuk mengukur performanya menggunakan tiga metrik: Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), dan R-squared (R^2). Fungsi objektif (fitness function) yang digunakan untuk proses optimasi dirumuskan sebagai berikut:

$$f(x) = \alpha \times (MAPE + RMSE - R^2) + (1 - \alpha) \times \frac{N_{selected}}{N_{features}} \quad (11)$$

Keterangan :

α = sebuah parameter yang mengontrol keseimbangan antara performa prediksi dan jumlah fitur,

MAPE = mengukur kesalahan relatif rata-rata prediksi,

RMSE = mengukur kesalahan absolut dalam satuan asli target,

R^2 = menunjukkan koefisien determinasi,

$N_{selected}$ = jumlah fitur terpilih,

$N_{features}$ = jumlah total fitur awal.

3.3. Optimasi Model

Optimasi model berguna untuk meningkatkan performa prediksi. Model yang digunakan ada SVM dengan metode regresi berbasis margin yang sensitive terhadap *hyperparameters*. Oleh karena itu, pemilihan nilai parameter yang optimal menjadi krusial dalam mencapai kinerja model yang maksimal. Tiga parameter utama dari SVM yang dioptomasi beserta rentang nilainya ditunjukkan pada Tabel 2 mencakup regularisasi (C), epsilon (ϵ) (*margin error*), dan gamma (γ) (kernel

coefficient) sangat mempengaruhi kemampuan model dalam menyesuaikan data sehingga pemilihan yang tepat menjadi krusial. Penggunaan algoritma GWO yang dapat secara efektif menjelajahi ruang parameter untuk menemukan kombinasi terbaik. Dalam pendekatan ini untuk setiap solusi GWO direpresentasikan sebagai vektor kontinu yang terdiri atas komponen seleksi fitur dan parameter SVM. Evaluasi tiap solusi dilakukan dengan melatih SVM pada subset fitur dan parameter tersebut dan kinerjanya diukur menggunakan tiga metrik utama regresi, yaitu: MAPE, RMSE, dan R-squared (R^2). Pendekatan ini memastikan bahwa model yang dihasilkan memiliki kombinasi terbaik antara akurasi dan efisiensi.

Tabel 2. Rentang Nilai Parameter

| Parameter | Rentang Nilai | Keterangan |
|--------------|--------------------------------------|--|
| Regularisasi | [0.001, 0.01, 0.1, 1, 10, 100, 1000] | Mengontrol keseimbangan antara margin dan kesalahan |
| Epsilon | [0.001, 0.01, 0.1, 0.5, 1.0] | Toleransi deviasi terhadap target tanpa penalti |
| Gamma | [0.0001, 0.001, 0.01, 0.1, 1, 10] | Pengaruh dari setiap data point dalam ruang fitur kernel RBF |

3.4. Validasi Model

Validasi model bertujuan untuk mengevaluasi kinerja model regresi dalam memprediksi nilai target berdasarkan fitur-fitur yang telah dipilih dan dioptimasi. Penelitian ini menggunakan tiga metrik evaluasi utama, yaitu Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE), dan R-squared (R^2) (Coefficient of Determination). Ketiga metrik ini dipilih karena mampu memberikan gambaran menyeluruh mengenai akurasi dan kesalahan dalam prediksi model regresi.

MAPE digunakan untuk mengukur rata-rata presentase kesalahan antara nilai actual dan nilai prediksi, dengan y_i sebagai nilai aktual, \hat{y}_i sebagai nilai prediksi, dan n sebagai jumlah data. Rumus MAPE dituliskan sebagai berikut:

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (12)$$

RMSE digunakan untuk mengukur akar dari rata-rata kuadrat selisih antara nilai actual dan prediksi yang lebih sensitif terhadap kesalahan besar. Rumus RMSE dituliskan sebagai berikut:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (13)$$

R^2 digunakan untuk mengukur sejauh mana variasi nilai target dapat dijelaskan oleh model. Nilai R^2 berkisar antara 0 hingga 1, semakin tinggi nilai mendekati angka satu maka kualitas prediksi semakin baik. Metrik ini menggunakan \bar{y} sebagai rata-rata dari nilai aktual dalam pengukurannya. Rumus R^2 dituliskan sebagai berikut:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (14)$$

DAFTAR PUSTAKA

- [1] R. L. Siegel, K. D. Miller, N. S. Wagle, and A. Jemal, "Cancer statistics, 2023," *CA. Cancer J. Clin.*, vol. 73, no. 1, pp. 17–48, Jan. 2023, doi: 10.3322/caac.21763.
- [2] I. W. A. S. Putra and I. Kuniawan, *2024 International Conference on Artificial Intelligence, Blockchain, Cloud Computing, and Data Analytics (ICoABCD)*. IEEE, 2024.
- [3] F. Alharbi and A. Vakanski, "Machine Learning Methods for Cancer Classification Using Gene Expression Data: A Review," Feb. 01, 2023, *MDPI*. doi: 10.3390/bioengineering10020173.
- [4] X. Wang, A. M. Bove, G. Simone, and B. Ma, "Molecular Bases of VEGFR-2-Mediated Physiological Function and Pathological Role," Nov. 16, 2020, *Frontiers Media S.A.* doi: 10.3389/fcell.2020.599281.
- [5] R. G. Yousef *et al.*, "(E)-N-(3-(1-(2-(4-(2,2,2-Trifluoroacetamido)benzoyl)hydrazono)ethyl)phenyl)nicotinamide: A Novel Pyridine Derivative for Inhibiting Vascular Endothelial Growth Factor Receptor-2: Synthesis, Computational, and Anticancer Studies," *Molecules*, vol. 27, no. 22, Nov. 2022, doi: 10.3390/molecules27227719.
- [6] E. B. Elkaeed *et al.*, "Discovery of New VEGFR-2 Inhibitors: Design, Synthesis, Anti-Proliferative Evaluation, Docking, and MD Simulation Studies," *Molecules*, vol. 27, no. 19, Oct. 2022, doi: 10.3390/molecules27196203.
- [7] L. Wang, W. Q. Liu, S. Broussy, B. Han, and H. Fang, "Recent advances of anti-angiogenic inhibitors targeting VEGF/VEGFR axis," 2023, *Frontiers Media SA*. doi: 10.3389/fphar.2023.1307860.
- [8] P. J. Chaudhari, A. R. Nemade, and A. A. Shirkhedkar, "Recent updates on potential of VEGFR-2 small-molecule inhibitors as anticancer agents," Oct. 22, 2024, *Royal Society of Chemistry*. doi: 10.1039/d4ra05244g.
- [9] H. Ding, F. Xing, L. Zou, and L. Zhao, "QSAR analysis of VEGFR-2 inhibitors based on machine learning, Topomer CoMFA and molecule docking," *BMC Chem.*, vol. 18, no. 1, pp. 1–17, 2024, doi: 10.1186/s13065-

024-01165-8.

- [10] F. Masoomi Sefiddashti, S. Asadpour, H. Haddadi, and S. Ghanavati Nasab, "QSAR analysis of pyrimidine derivatives as VEGFR-2 receptor inhibitors to inhibit cancer using multiple linear regression and artificial neural network," *Res. Pharm. Sci.*, vol. 16, no. 6, pp. 596–611, Dec. 2021, doi: 10.4103/1735-5362.327506.
- [11] V. Napolitano *et al.*, "Structure-based design, synthesis and evaluation of a novel family of PEX5-PEX14 interaction inhibitors against Trypanosoma," *Eur. J. Med. Chem.*, vol. 243, no. September, 2022, doi: 10.1016/j.ejmech.2022.114778.
- [12] A. Salimi, J. H. Lim, J. H. Jang, and J. Y. Lee, "The use of machine learning modeling, virtual screening, molecular docking, and molecular dynamics simulations to identify potential VEGFR2 kinase inhibitors," *Sci. Rep.*, vol. 12, no. 1, pp. 1–14, 2022, doi: 10.1038/s41598-022-22992-6.
- [13] N. Arabi, M. R. Torabi, and F. Ghasemi, "Application of Support Vector Machine Classification Model to Identification of Vascular Endothelial Growth Factor Receptor Inhibitors," *Adv. Biomed. Res.*, vol. 13, no. 1, 2024, doi: 10.4103/abr.abr_179_23.
- [14] L. Chen, H. Fan, and H. Zhu, "Multi-Objective Optimization of Cancer Treatment Using the Multi-Objective Grey Wolf Optimizer (MOGWO)," *Int. J. Performability Eng.*, vol. 19, no. 7, pp. 481–490, 2023, doi: 10.23940/ijpe.23.07.p7.481490.
- [15] R. G. Yousef *et al.*, "Anti-cancer and immunomodulatory evaluation of new nicotinamide derivatives as potential VEGFR-2 inhibitors and apoptosis inducers: in vitro and in silico studies," *J. Enzyme Inhib. Med. Chem.*, vol. 37, no. 1, pp. 2206–2222, 2022, doi: 10.1080/14756366.2022.2110868.
- [16] Y. Zhang, Y. Gao, L. Huang, and X. Xie, "An Improved Grey Wolf Optimizer Based on Attention Mechanism for Solving Engineering Design Problems," *Symmetry (Basel)*, vol. 17, no. 1, pp. 1–52, 2025, doi: 10.3390/sym17010050.
- [17] R. Liu *et al.*, "Hybrids of support vector regression with grey wolf optimizer and firefly algorithm for spatial prediction of landslide susceptibility,"

Remote Sens., vol. 13, no. 24, 2021, doi: 10.3390/rs13244966.

LAMPIRAN

Data-data pendukung, *curriculum vitae*(CV) untuk pembimbing dari luar Universitas Telkom, dsb.