PREDIKSI HARGA SAHAM ANTM MENGGUNAKAN *HYBRID SUPPORT VECTOR REGRESSION–FIREFLY ALGORITHM*

Zidane Arryo[1], Taghfirul Azhima Yoga Siswa[2*], Rofilde Hasudungan[3]

[123,]Program Studi Teknik Informatika, Universitas Muhammadiyah Kalimantan Timur, Samarinda. Indonesia

Jl. Ir. H. Juanda No.15, Sidodadi, Kec. Samarinda Ulu, Kota Samarinda, Kalimantan Timur 75124
[1]2211102441095@umkt.ac.id, [2*]tay758@umkt.ac.id, [3] rh219@umkt.ac.id

Abstract

The fluctuating stock prices create uncertainty for investors, requiring a predictive method capable of handling nonlinear patterns in time series data. This study evaluates the performance of the hybrid Support Vector Regression–Firefly Algorithm (SVR-FA) model in predicting the stock prices of PT Aneka Tambang Tbk (ANTM.JK), compared to the standard SVR model. The data used consists of daily stock prices from September 2020 to September 2025. The research process includes data preprocessing, lag feature creation, normalization, 80:20 data splitting, and time series cross-validation. SVR parameter optimization is performed using the Firefly Algorithm (FA) to find the best parameters. The results show that SVR-FA achieved RMSE of 76.96, MAE of 54.42, and MAPE of 2.55%, outperforming the SVR model, which produced RMSE of 93.17, MAE of 63.30, and MAPE of 2.77%. These results show that optimization using FA is able to reduce prediction errors and improve the performance of the SVR model.

Keywords : *prediction, stock, support vector regression, firefly algorithm, machine learning.*

Abstrak

Harga saham yang fluktuatif memunculkan ketidakpastian bagi investor, sehingga dibutuhkan metode prediksi yang dapat mengatasi pola nonlinier dalam data deret waktu. Penelitian ini mengevaluasi kinerja model *hybrid Support Vector Regression–Firefly Algorithm* (SVR-FA) dalam memprediksi harga saham PT Aneka Tambang Tbk (ANTM.JK), dibandingkan dengan model SVR standar. Data yang digunakan adalah harga saham harian dari September 2020 hingga September 2025. Proses penelitian meliputi preprocessing, pembentukan fitur *lag*, normalisasi, pembagian data 80:20, dan *cross-validation time series*. Optimasi parameter SVR menggunakan *Firefly Algorithm* (FA) untuk mencari parameter terbaik. Hasil penelitian menunjukkan bahwa SVR-FA menghasilkan RMSE 76.96, MAE 54.42, dan MAPE 2.55%, lebih baik dibandingkan model SVR yang menghasilkan RMSE 93.17, MAE 63.30, dan MAPE 2.77%. Hasil tersebut menunjukkan bahwa optimasi menggunakan FA mampu menurunkan kesalahan prediksi dan meningkatkan kinerja model SVR.

Kata kunci : *prediksi, saham, support vector regression, firefly algorithm, pembelajaran mesin*.

1. PENDAHULUAN

Saham merupakan surat berharga yang menjadi tanda bukti kepemilikan atas suatu perusahaan, saham dapat dipahami sebagai instrumen sekuritas yang menunjukkan bahwa pemegang saham bisa memiliki hak kepemilikan terhadap aset perusahaan [1]. Menurut data Bursa Efek Indonesia, hingga 13 April 2020 tercatat 686 perusahaan terdaftar, menjadikan saham sebagai instrumen investasi yang diminati di Indonesia. Saham Aneka Tambang (ANTM) menunjukkan kinerja unggul dengan penguatan mingguan 3.36% dan bulanan 11.18%, melampaui beberapa saham unggulan lain. Periode 2020–2024 mencatat kenaikan harga ANTM sebesar 351% serta pertumbuhan laba bersih 25%, memperkuat prospeknya sebagai instrumen investasi yang menarik [2].

Cara memprediksi pasar saham tidak lagi hanya mengandalkan metode tradisional seperti membaca pola grafik harga dan menilai kondisi keuangan perusahaan, kini metode tersebut mulai berkembang dengan memanfaatkan teknologi *modern* seperti *Machine Learning* [3]. *Machine Learning* merupakan teknik yang digunakan oleh sistem untuk belajar dari data pelatihan yang sesuai dengan konteks tertentu, lalu secara otomatis membangun model analisis dan mengenali pola dalam data untuk menghasilkan prediksi maupun keputusan [4]. Beberapa penelitian sudah menerapkan algoritma *ML* dalam memprediksi harga saham, di antaranya *K-Nearest Neighbors (KNN)*, *Linear Regression*, *Support Vector Regression (SVR)*, *Decision Tree Regression*, serta *Long Short-Term Memory (LSTM)* [5].

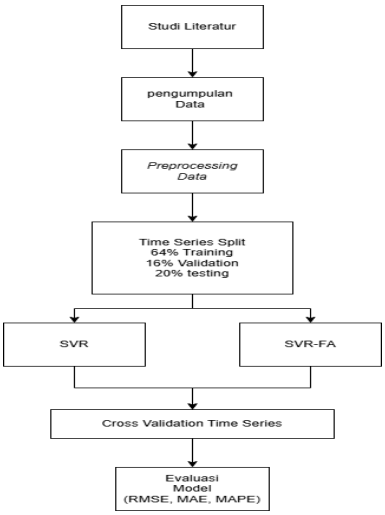
Studi terdahulu yang dilakukan tentang prediksi saham *Moroccan Stock Index 20* menunjukkan hasil penelitian bahwa *SVR* lebih unggul dengan *MSE* 15.941 dan *RMSE* 3.993, dibandingkan *XGBoost* yang menghasilkan *MSE* 182.651 dan *RMSE* 13.515 [6]. Sementara itu, penelitian lain pada *Pakistan Stock Exchange* menunjukkan bahwa *SVR* memiliki performa terbaik dengan *MAE* 1401.253, *RMSE* 1562.116, *MAPE* 2.489, dan koefisien determinasi (R^2) 0.976, model ini mengungguli *Decision Tree* dengan *MAE* 10790.75, *RMSE* 13714.49, *MAPE* 17.992, dan R^2 0.843 *Random Forest* dengan *MAE* 12415.67, *RMSE* 15483.47, *MAPE* 20.887, dan R^2 0.864 serta *KNN* dengan *MAE* 11251.1, *RMSE* 13955.62, *MAPE* 18.924, dan R^2 0.908 [7].

Support Vector Regression (SVR) merupakan turunan dari *Support Vector Machine (SVM)* yang dirancang untuk menangani permasalahan regresi dengan output berupa nilai kontinu[8]. Keunggulan *SVR* terletak pada kemampuannya dalam menangani data yang kompleks, khususnya ketika hubungan antar variabel bersifat *nonlinier* dan sulit dimodelkan dengan pendekatan konvensional [9]. Namun, penggunaan *SVR* dengan parameter *default* memiliki keterbatasan dalam memprediksi harga yang tidak menentu, hal ini dibuktikan oleh penelitian prediksi harga saham *bitcoin* yang menunjukkan bahwa *SVR* memberikan hasil kurang optimal, dengan *MAE* berkisar 6.500–6.900 dan *MSE* melebihi 83 juta. Sebaliknya, algoritma *LSTM* menghasilkan prediksi yang lebih presisi, ditandai dengan *MAE* 1.400–1.700 serta *MSE* hanya 3–5 juta, oleh karena itu, peneliti merekomendasikan penerapan optimasi *hyperparameter* untuk meningkatkan kinerja model prediksi [10].

Beberapa metode optimasi *metaheuristic* yang dapat digunakan dalam meningkatkan kinerja model *SVR* di antaranya adalah *Genetic Algorithm (GA)*, *Particle Swarm Optimization (PSO)*, *Artificial Bee Colony (ABC)*, *Firefly Algorithm (FA)*, dan *Grey Wolf Optimizer (GWO)* [11]. *FA* merupakan salah satu metode optimasi yang terinspirasi dari perilaku kunang-kunang, algoritma ini bekerja dengan meniru cara kunang-kunang saling menarik satu sama lain menggunakan cahaya mereka untuk menemukan solusi terbaik dari suatu permasalahan [12].

Penelitian tentang prediksi konsumsi energi di instalasi pengolahan air limbah menggunakan *SVR–FA* memperoleh hasil lebih baik dibandingkan *SVR* tanpa optimasi, dengan nilai *RMSE* 410.6 *kWh*, dan *MAPE* 6.43%, sementara itu *SVR* tanpa optimasi menghasilkan *MAPE* 6.80% dan *RMSE* 436.5 *kWh* [13]. Namun, penerapan *hybrid SVR-FA* untuk memprediksi data deret waktu finansial yang sangat fluktuatif, seperti saham sektor pertambangan di Indonesia, belum banyak dieksplorasi secara komprehensif menggunakan validasi berbasis urutan waktu. Penelitian ini berfokus untuk menganalisis penerapan model *hybrid SVR-FA* dalam memprediksi harga saham PT Aneka Tambang Tbk (ANTM) dengan mempertimbangkan urutan temporal data melalui *time series cross-validation*. Kontribusi utama penelitian ini terletak pada evaluasi empiris mengenai sejauh mana optimasi *Firefly Algorithm* mampu meningkatkan akurasi dan menurunkan tingkat kesalahan *SVR* berdasarkan metrik evaluasi *RMSE*, *MAE*, dan *MAPE*.

2. METODOLOGI PENELITIAN



Gambar 1. Diagram Alur Penelitian

Berdasarkan Gambar 1, penelitian ini terdiri dari beberapa tahap yang dimulai dari studi literatur, pengumpulan data, dan *preprocessing data*. Setelah tahap *preprocessing*, data dibagi menggunakan metode *time series split* dengan proporsi 64% *data training*, 16% *data validation*, dan 20% *data testing*. *Data training* dan *validation* kemudian digunakan dalam proses pelatihan dan optimasi model menggunakan algoritma *SVR* serta *SVR* yang dioptimasi dengan *FA*. Selanjutnya, *Time series cross-validation* digunakan pada *data training* dan *data validation* untuk proses *tuning* parameter serta evaluasi kestabilan model. Setelah parameter optimal diperoleh. Tahap akhir penelitian adalah evaluasi kinerja model dengan metrik *RMSE*, *MAE*, dan *MAPE*.

2.1 Pengumpulan Data

Data yang digunakan dalam penelitian ini berasal dari data historis saham perusahaan PT Aneka Tambang Tbk (ANTM), yang mencakup informasi (*open*) harga pembukaan, (*high*) harga tertinggi, (*low*) harga terendah, (*close*) harga terakhir, harga perubahan% (*change* **%**), serta *volume* transaksi perdagangan saham. Korpus data dihimpun dari portal Investing.com dengan cakupan observasi selama lima tahun ke belakang. Untuk kebutuhan analisis dan visualisasi visual, penelitian ini berfokus pada empat variabel utama: Open, High, Low, dan Close. Sementara itu, fitur pelengkap seperti Date, Volume, serta persentase perubahan dieleminasi pada tahap seleksi data guna menjaga relevansi input model. Cuplikan sampel data tersebut disajikan pada Tabel I.

TABEL I. DATA PENELITIAN

Close	***Open***	***High***	***Low***
3.390	3.530	3.570	3.380
3.510	3.580	3.600	3.470
3.660	3.680	3.930	3.590
3.610	3.470	3.710	3.450
3.390	3.470	3.470	3.370
3.480	3.520	3.580	3.360
3.470	3.300	3.480	3.290

Pada Tabel I memperlihatkan data historis saham yang menjadi dasar analisis dalam penelitian ini. Data yang digunakan berupa deret waktu harian saham PT Aneka Tambang Tbk, yang diperoleh melalui situs *Investing.com* untuk periode 11 September 2020 hingga 11 September 2025. Secara keseluruhan terkumpul 1.202 data, yang mencakup informasi harga terakhir (*close*), harga pembukaan (*open*), harga tertinggi (*high*), dan harga terendah (*low*).

2.2 *Preprocessing* Data

Dilakukan tahap *preprocessing* meliputi *data selection*, *data transformation, lag feature,* normalisasi data, denormalisasi data, serta *cross validation* agar data dapat digunakan dalam proses analisis dan pemodelan [14].

Data Selection adalah prosedur untuk memilih bagian data yang paling konsisten dengan model yang digunakan, agar prediksi atau inferensi tetap akurat meskipun data keseluruhan tidak sepenuhnya cocok dengan model [15]. Fokus visualisasi diarahkan sepenuhnya pada empat variabel harga, yakni Open, High, Low, dan Close. Konsekuensinya, atribut pelengkap seperti Date, Volume, serta tingkat perubahan harga dieliminasi dari rangkaian analisis..

Data Transformation adalah proses mengubah dan menyesuaikan data dari berbagai sumber agar berada dalam format yang dibutuhkan, sehingga dapat digunakan untuk analisis, proses ini meliputi pengubahan tipe data, skala, dan standarisasi [16].

Pemilihan *lag-time* memegang peranan krusial dalam pemodelan deret waktu; ketepatan menentukan rentang observasi lampau memastikan data input tetap berkorelasi kuat dengan target proyeksi serta mendongkrak ketepatan prediksi [17]. Dalam rancangan eksperimen ini, parameter lag yang ditetapkan dibatasi sebesar satu periode pengamatan.

Normalisasi adalah proses penting dalam *data preprocessing* yang bertujuan untuk menyetarakan skala data, data dengan rentang nilai yang terlalu besar dapat memengaruhi kinerja algoritma, sehingga diperlukan transformasi agar seluruh atribut berada dalam skala yang sebanding tanpa mengubah makna informasi yang terkandung metode normalisai yang dilakukan dalam penelitian ini adalah *Min-Max* [18].

Prosedur denormalisasi berfungsi mengonversi kembali data yang telah diskalakan ke dalam rentang metrik aslinya. Langkah ini krusial agar interpretasi serta evaluasi luaran model dapat diukur berdasarkan besaran riil yang sebenarnya [19].

Cross-validation pada data *time series* merupakan metode evaluasi yang digunakan untuk menilai kinerja model prediksi dengan tetap mempertahankan urutan waktu dalam proses pemisahan data. Berbeda dengan *cross-validation* pada *data non-time series* yang dapat

dilakukan secara acak, pada data *time series* urutan temporal harus dijaga karena observasi yang berdekatan cenderung saling berkaitan serta membentuk pola atau tren tertentu. Oleh karena itu, data latih selalu berasal dari periode waktu sebelum data uji agar tidak terjadi penggunaan informasi masa depan dalam proses pelatihan model [20].

2.3 Algoritma Support Vector Regression

Metode *Support Vector Regression (SVR)* dirancang untuk menyelesaikan permasalahan yang bersifat *non-linear* dan telah terbukti efektif dalam berbagai studi prediksi deret waktu *(time series)*. Oleh karena itu, *SVR* dikategorikan sebagai salah satu algoritma yang dapat diterapkan pada tugas klasifikasi, identifikasi, regresi, maupun analisis deret waktu. Dengan kemampuannya dalam memodelkan hubungan regresi pada data yang tidak linear [21]. Perhitungan linear *SVR* dapat dilihat pada persamaan (1).

$$f(x) = w^T x + b \quad (1)$$

Keterangan:
$f(x)$ = Nilai prediksi atau *output* dari model untuk *input x.*
w = Vektor Bobot.
x = Vektor *Input.*
$w^T x$ = *Dot product* antara bobot dan *input*
b = Bias atau *intercept,* nilai tambahan yang menggeser nilai prediksi.

2.4 Optimasi *Firefly Algorithm*

Firefly Algorithm bekerja dengan prinsip bahwa kunang-kunang dengan intensitas cahaya yang lebih tinggi akan menarik kunang-kunang lain yang lebih redup untuk mendekat, sehingga posisi dalam ruang solusi diperbarui secara iteratif hingga mencapai solusi optimal, keunggulan *FA* terletak pada modelnya yang sederhana, konsep yang jelas, serta efektivitasnya dalam menyelesaikan berbagai permasalahan optimasi global [22]. Untuk menghitung intensitas cahaya *FA* dapat dilihat pada persamaan 2.

$$I(r) = I_0 \cdot e^{-\gamma r_{ij}^2} vv \quad (2)$$

$I(r)$ menyatakan intensitas cahaya pada jarak r, I_0 adalah intensitas cahaya awal atau intensitas maksimum, *e* merupakan bilangan eksponensial natural, γ adalah koefisien penyerapan atau faktor redaman, r_{ij} menyatakan jarak antara kunang-kunang ke-i dan ke-j, serta ν merupakan parameter yang mengatur laju penurunan intensitas terhadap jarak.

Untuk menghitung jarak antar FA dapat dilihat pada persamaan 3 berikut :

$$r_{ij} = |x_i - x_j| = \sqrt{\sum_{k=1}^{d} (x_{i,k} - x_{j,k})^2} \quad (3)$$

r_{ij} menyatakan jarak *Euclidean* antara individu atau solusi ke-i dan ke-j, x_i dan x_j masing-masing adalah vektor posisi solusi ke-i dan ke-j, d menyatakan jumlah dimensi atau parameter yang digunakan, $x_{i,k}$ dan $x_{j,k}$ adalah nilai komponen ke-k dari vektor x_i dan x_j, serta $|x_i - x_j|$ menyatakan besar jarak antara kedua vektor tersebut.

Untuk menghitung daya tarik *firefly* dapat dilihat pada persamaan 4 :

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

β(r) menyatakan tingkat daya tarik *(attractiveness)* pada jarak *r*, β_0 adalah nilai daya tarik awal pada jarak nol, *e* merupakan bilangan eksponensial natural, *γ* adalah parameter penurunan daya tarik, dan r_{ij} menyatakan jarak antara kunang-kunang ke-*i* dan ke-*j.*

Untuk menghitung *update* posisi *firefly* dapat dilihat pada persamaan 5 :

$$x_i^{t+1} = x_i^t + \beta_0 e^{-\gamma r_{ij}^2} \cdot (x_j^t - x_i^t) + \alpha \cdot (Rnd - 0.5) \quad (5)$$

x_i^{t+1} menyatakan posisi kunang-kunang ke-i pada iterasi ke-t+1, x_i^t adalah posisi kunang-kunang ke-i pada iterasi ke-t, β_0 merupakan daya tarik awal, e adalah bilangan eksponensial natural, γ menyatakan parameter redaman daya tarik, r_{ij} adalah jarak antara kunang-kunang ke-i dan ke-j, $(x_j^t - x_i^t)$ menunjukkan arah perpindahan menuju kunang-kunang yang lebih terang, α adalah parameter pengendali komponen acak, dan *Rnd* merupakan bilangan acak dengan rentang 0 dan 1.

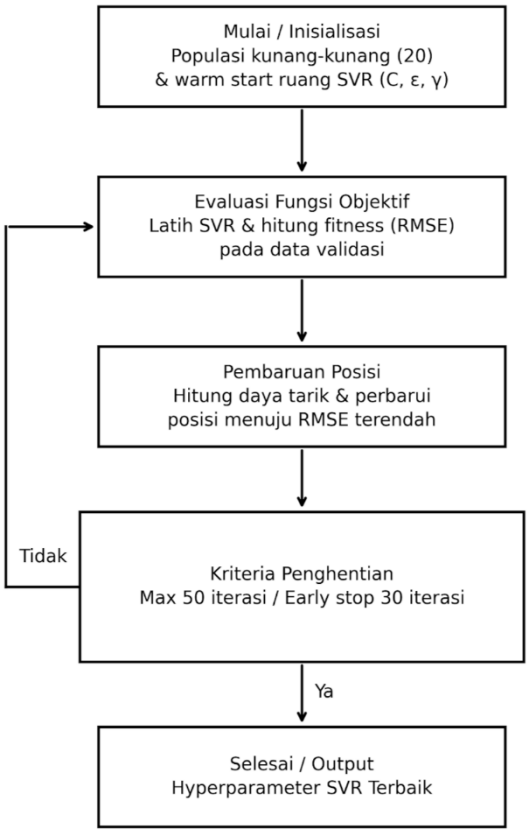
FA diintegrasikan dengan *SVR* untuk mencari kombinasi *hyperparameter* yang paling optimal secara otomatis. Setiap populasi kunang-kunang dalam algoritma ini direpresentasikan sebagai vektor tiga dimensi yang berisi parameter model *SVR*, yaitu nilai penalti regulerisasi *(C)*, batas margin kesalahan *(epsilon),* dan parameter kernel *(gamma).* Ruang pencarian *(search space)* dibatasi dengan rentang nilai yang spesifik untuk mencegah terjadinya *overfitting*, yakni parameter *C* dengan rentang antara 50 hingga 500, *epsilon* antara 0.00001 hingga 0.01, dan *gamma* antara 0.001 hingga 0.1. Selain

penyebaran acak, proses inisialisasi awal pada algoritma ini juga menerapkan metode *warm start*, di mana salah satu kunang-kunang secara eksplisit ditempatkan pada titik koordinat parameter terbaik yang diperoleh dari hasil uji iterasi parameter sebelumnya. Langkah ini bertujuan untuk mempercepat proses konvergensi algoritma menuju solusi yang paling optimal.

Kualitas atau intensitas cahaya *(fitness)* dari setiap kunang-kunang diukur berdasarkan fungsi objektif, yaitu minimasi tingkat kesalahan prediksi. Untuk setiap iterasi, kombinasi parameter *(C, epsilon, gamma)* yang direpresentasikan oleh kunang-kunang akan diterapkan ke dalam model *SVR*. Model tersebut kemudian dilatih menggunakan *data training* dan diuji kinerjanya menggunakan *data validation* melalui mekanisme *time series holdout*. Performa dari model dievaluasi secara spesifik menggunakan metrik *Root Mean Squared Error (RMSE)*. Kunang-kunang yang menghasilkan nilai *RMSE* terendah diposisikan sebagai individu dengan intensitas cahaya paling terang. Berdasarkan prinsip dasar *FA*, kunang-kunang dengan tingkat kesalahan prediksi yang lebih tinggi (redup) akan bergerak mendekati posisi parameter kunang-kunang dengan nilai *RMSE* terendah (terang).

Proses pembaruan posisi kunang-kunang ini akan terus dieksekusi secara berulang hingga sistem memenuhi salah satu dari dua kriteria penghentian *(stopping criteria)*. Kriteria utama adalah batas iterasi maksimal *(maximum iterations)*, di mana proses komputasi algoritma ditetapkan akan berhenti setelah menyelesaikan 50 iterasi optimasi. Kriteria kedua adalah mekanisme penghentian dini *(early stopping)*, yang dirancang untuk menghemat beban dan waktu komputasi apabila model telah mencapai titik konvergensi. Melalui kriteria ini, algoritma akan dihentikan secara otomatis apabila nilai *RMSE* terbaik tidak mengalami penurunan signifikansi atau perbaikan minimal sebesar 1e-6 selama 30 iterasi secara berturut-turut *(early stopping patience = 30)*. Posisi parameter dari kunang-kunang terbaik pada saat kriteria penghentian terpenuhi akan ditetapkan sebagai parameter final untuk pengujian model *hybrid SVR-FA*.

Secara keseluruhan, tahapan proses pencarian parameter optimal SVR menggunakan Firefly Algorithm dapat dilihat pada Gambar 2.



Gambar 2. Diagram Alur Optimasi *SVR* menggunakan *FA*

2.5 Evaluasi Model

Pada penelitian ini beberapa evaluasi yang digunakan meliputi *Mean Squared Error (MSE)*, *Mean Absolute Error (MAE)*, *Mean Absolute Percentage Error (MAPE)*, *Root Mean Squared Error (RMSE)* evaluasi ini digunakan untuk mengevaluasi dan membandingkan tingkat keakuratan hasil prediksi dari model yang dikembangkan.

a. *Root Mean Squared Error (RMSE)*

Root Mean Square Error (RMSE) adalah ukuran standar dalam analisis regresi yang menghitung rata-rata besar kesalahan prediksi. Jadi, *RMSE* menunjukkan seberapa jauh hasil prediksi dari nilai sebenarnya, dengan memberi bobot lebih besar pada kesalahan yang besar [23]. Perhitungan *RMSE* dapat dilihat pada persamaan 6.

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2} \qquad (6)$$

n menyatakan jumlah total data pengamatan, y_t adalah nilai aktual pada waktu t,

$\hat{y}_t$ merupakan nilai hasil prediksi model pada waktu t, dan $\sum$ menyatakan operasi penjumlahan dari t = 1 hingga n.

b. *Mean Absolute Error (MAE)*

Mean Absolute Error (MAE) adalah ukuran statistik yang menghitung rata-rata selisih absolut antara nilai prediksi dan nilai sebenarnya, jadi *MAE* menggambarkan seberapa banyak rata-rata kesalahan tanpa memperhatikan arahnya (positif/negatif) [23]. Perhitungan *MAE* dapat dilihat pada persamaan 7.

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|y_t - \hat{y}_t| \quad (7)$$

n menyatakan jumlah total data pengamatan, y_t adalah nilai aktual pada waktu t, $\hat{y}_t$ merupakan nilai hasil prediksi model pada waktu t, dan $|y_t - \hat{y}_t|$ menyatakan besar selisih absolut antara nilai aktual dan nilai prediksi.

c. *Mean Absolute Percentage Error (MAPE)*

Mean Absolute Percentage Error (MAPE) dihitung dengan cara membagi kesalahan absolut pada tiap periode dengan nilai aktual pada periode tersebut, lalu mengambil rata-rata dari persentase kesalahan absolut itu [24]. Perhitungan *MAPE* dapat dilihat pada persamaan 8.

$$MAPE = \frac{100\%}{n}\sum_{t=1}^{n}\left|\frac{y_t - \hat{y}_t}{y_t}\right| \quad (8)$$

n menyatakan jumlah total data pengamatan, y_t adalah nilai aktual pada waktu *t*, $\hat{y}_t$ merupakan nilai hasil prediksi model pada waktu *t*, $|y_t - \hat{y}_t|$ menyatakan besar selisih absolut antara nilai aktual dan nilai prediksi, serta *100%* digunakan untuk mengonversi nilai kesalahan ke dalam bentuk persentase.

3. HASIL DAN PEMBAHASAN

3.1 Pengujian Cross Validation

Cross-Validation diterapkan menggunakan metode *time series cross-validation* dengan menjaga urutan temporal data, di mana model dilatih pada data periode sebelumnya dan divalidasi pada periode setelahnya untuk memperoleh evaluasi kinerja yang stabil sebelum pengujian akhir. Hasil pengujian dapat dilihat pada Tabel II.

TABEL II. PENGUJIAN CROSS VALIDATION SVR

Fold	Data Latih	Data Validasi	*RMSE*	*MAE*	*MAPE%*
1	160	160	54.28	39.95	1.78%
2	320	160	81.88	56.27	2.48%
3	480	160	73.91	63.38	3.14%
4	640	160	45.44	35.49	2.21%
5	800	160	55.89	46.49	3.36%

Berdasarkan Tabel II, terlihat hasil implementasi *time series cross-validation* dengan skema *expanding window* pada model *SVR* standar, di mana ukuran data latih bertambah secara bertahap pada setiap *fold* sementara ukuran data validasi tetap konstan sebesar 160 data. *Fold* pertama menggunakan 160 data latih dan menghasilkan nilai *RMSE* 54.28, *MAE* 39.95, serta *MAPE* 1.78%. Pada *fold* kedua dan ketiga, seiring dengan peningkatan jumlah data latih (320 dan 480 data), terjadi kenaikan tingkat kesalahan, dengan *fold* kedua mencapai *RMSE* 81.88 dan *fold* ketiga mencatatkan *MAE* tertinggi sebesar 63.38. Hal ini menunjukkan bahwa model *SVR* standar mengalami fluktuasi dan kesulitan dalam menangkap variasi pola data secara optimal pada fase awal perluasan data temporal. Meskipun kinerja model sempat membaik pada *fold* keempat dengan penurunan *RMSE* yang signifikan menjadi 45.44, kesalahan kembali meningkat pada *fold* kelima (*RMSE* 55.89 dan *MAPE* 3.36%). Tingkat kesalahan yang berfluktuasi lintas *fold* ini mengindikasikan bahwa penggunaan *SVR* tanpa optimasi heuristik memiliki kerentanan (sensitivitas) terhadap penambahan deret waktu, sehingga belum mampu memberikan stabilitas generalisasi yang sepenuhnya konsisten pada keseluruhan data latih.

3.2 Penerapan algoritma *SVR*

a. Pengujian Iterasi

Pengujian iterasi dilakukan untuk menganalisis pengaruh jumlah iterasi terhadap konvergensi dan stabilitas kinerja model. Beberapa nilai iterasi diuji untuk memperoleh keseimbangan antara akurasi prediksi dan efisiensi komputasi. Hasil pengujian dapat dilihat pada Tabel III.

TABEL III. PENGUJIAN ITERASI

Iterasi	*RMSE*	*MAE*	*MAPE*
1000	34.41	25.30	1.67%
600	34.64	25.52	1.69%
900	34.97	25.60	1.69%
800	35.30	25.81	1.70%

700	36.21	26.70	1.76%
500	36.53	26.70	1.74%
400	42.33	32.55	2.10%
300	52.42	41.70	2.65%
200	53.24	44.55	2.91%
100	169.99	166.40	11.06%

Peningkatan jumlah iterasi sebanding dengan penurunan *RMSE*, *MAE*, dan *MAPE*, menunjukkan proses optimasi yang semakin stabil. Pada iterasi rendah (100–300), kesalahan masih tinggi. Namun, seiring bertambahnya iterasi (500–800), metrik kesalahan mulai menurun. Iterasi 900–1000 menghasilkan error paling kecil dan stabil, menandakan bahwa iterasi lebih banyak meningkatkan kinerja, meski dengan laju perbaikan yang menurun

b. Pengujian Parameter C

Pengujian parameter *C* pada model *SVR* bertujuan untuk menentukan nilai terbaik dari parameter regulasi, yang mengontrol *trade-off* antara kompleksitas model dan kesalahan pada data pelatihan. Hasil pengujian dapat dilihat pada Tabel IV.

Tabel IV. PENGUJIAN PARAMETER C

Range C	Nilai Terbaik	*RMSE*	*MAE*	*MAPE*
100-200	170.99	34.79	25.31	1.66%
200-300	240.07	89.29	80.75	5.57%
300-400	305.55	35.86	26.25	1.72%
400-500	475.87	111.60	105.67	7.08%
500-600	572.71	272.10	235.82	16.68%
600-700	617.78	159.92	137.98	9.76%
700-800	765.49	81.19	67.98	4.80%
800-900	868.78	155.99	147.78	10.05%
900-1000	982.48	37.45	28.97	1.95%

Berdasarkan hasil pengujian, Rentang 100–200 memberikan kinerja optimal dengan *RMSE* 34.79, *MAE* 25.31, dan *MAPE* 1.66%, mencerminkan kemampuan generalisasi model yang baik. Rentang 300–400 dan 900–1000 menghasilkan kesalahan rendah, sementara rentang 500–600 dan 800–900 menunjukkan peningkatan error yang cukup besar. Rentang 600–700 dan 700–800 memiliki kinerja menengah, namun tidak sebaik rentang *C* yang lebih kecil.

c. Pengujian Parameter *Epsilon*

Parameter epsilon pada model *SVR* mengontrol margin di sekitar fungsi regresi, menentukan seberapa besar kesalahan yang dianggap dapat diterima dalam model tanpa mempengaruhi proses pelatihan. Hasil pengujian dapat dilihat pada Tabel V.

TABEL V. PENGUJIAN PARAMETER EPSILON

Range Epsilon	Nilai Terbaik	*RMSE*	*MAE*	*MAPE*
0.0001-0.01	0.000755	34.48	25.46	1.67%
0.001-0.01	0.007221	34.67	25.80	1.70%
0.0001-0.001	0.000594	34.82	25.47	1.67%
0.001-0.1	0.024816	44.21	36.08	2.41%

Pengujian parameter epsilon pada model *SVR* menunjukkan bahwa rentang 0.0001–0.01 memberikan kinerja optimal dengan *RMSE* 34.48, *MAE* 25.46, dan *MAPE* 1.67%, mencerminkan kesalahan rendah dan kestabilan model. Rentang 0.001–0.01 dan 0.0001–0.001 menunjukkan performa serupa dengan selisih kesalahan kecil, sementara rentang 0.001–0.1 menghasilkan peningkatan kinerja yang lebih baik, menandakan nilai *epsilon* yang terlalu besar menurunkan sensitivitas model

d. Pengujian Parameter *Gamma*

Parameter *gamma* mengatur kekuatan pengaruh setiap titik data pada model, mempengaruhi bentuk fungsi *kernel* dan seberapa jauh pengaruh titik data terhadap titik lainnya dalam ruang fitur. Hasil pengujian dapat dilihat pada Tabel VI.

TABEL VI. PENGUJIAN PARAMETER *GAMMA*

Range Gamma	Nilai Terbaik	RMSE	MAE	MAPE
0.001-0.1	0.024816	34.36	25.12	1.65%
0.01-0.1	0.012422	34.46	25.46	1.67%
0.001-0.01	0.007221	35.08	26.27	1.73%
0.0001-0.01	0.000755	35.84	26.42	1.74%
0.0001-0.001	0.000594	36.17	26.65	1.75%

Rentang 0.001–0.1 memberikan kinerja paling optimal dengan *RMSE* 34.36, *MAE* 25.12, dan *MAPE* 1.65%, yang menandakan kemampuan model menangkap pola *nonlinier* secara efektif. Rentang 0.01–0.1 dan 0.001–0.01 masih menunjukkan performa yang baik dengan peningkatan kesalahan relatif kecil, sedangkan rentang 0.0001–0.01 dan 0.0001–0.001 menghasilkan *error* lebih tinggi, yang mengindikasikan bahwa nilai *gamma* terlalu kecil membuat model kurang responsif terhadap variasi data.

e. Parameter Terbaik *SVR*

Parameter terbaik *SVR* adalah kombinasi nilai Iterasi, *C*, *epsilon*, dan *gamma* yang menghasilkan kinerja model terbaik dengan kesalahan prediksi terendah berdasarkan metrik evaluasi seperti *RMSE, MAE*, dan *MAPE*.

TABEL VII. PARAMETER TERBAIK *SVR*

Iterasi	*C*	*Epsilon*	*Gamma*
1000	170.99	0.000754	0.024816

Berdasarkan Tabel VI, Konfigurasi parameter terbaik *SVR* yang terdiri atas nilai Iterasi sebesar 1000, *C* sebesar 170.99, *epsilon* sebesar 0.000754, dan *gamma* sebesar 0.024816, yang diperoleh berdasarkan hasil pengujian parameter yang telah dilakukan sebelumnya dengan mempertimbangkan nilai kesalahan prediksi terendah.

f. Evaluasi Model *SVR*

Evaluasi model *SVR* mengukur kinerja model dengan membandingkan hasil prediksi terhadap data aktual menggunakan metrik seperti *RMSE*, *MAE*, dan *MAPE* untuk menilai akurasi dan kesalahan prediksi. Hasil evaluasi dapat dilihat pada Tabel VIII.

TABEL VIII. EVALUASI MODEL SVR

RMSE	*MAE*	*MAPE*
93.17	63.30	2.77%

3.3 Penerapan Optimasi *FA*

Untuk meningkatkan kinerja model *SVR*, pada penelitian ini diterapkan *FA* sebagai metode optimasi parameter. *FA* digunakan untuk mencari kombinasi parameter *SVR* yang optimal, sehingga kesalahan prediksi dapat diminimalkan. Berikut penentuan parameter *FA* yang digunakan pada penelitian ini dapat dilihat pada tabel IX.

TABEL IX. PARAMETER *FA*

Parameter	Nilai Optimal
Beta0	0.8
Gamma_fa	2.0
Alpha	0.05
Alpha_decay	0.94
Beta_min	0.1
N_fireflies	20
Max Iterasi FA	50

Berdasarkan Tabel VIII, nilai parameter optimal yang digunakan pada algoritma *Firefly Algorithm* (FA) meliputi nilai *Beta0* sebesar 0,8, *Gamma_fa* sebesar 2,0, dan *Alpha* sebesar 0,05. Selanjutnya, nilai *Alpha_decay* ditetapkan sebesar 0,94 serta *Beta_min* sebesar 0,1. Pengujian ini menggunakan jumlah populasi (*N_fireflies*) sebanyak 20 kunang-kunang dengan batas *Max Iterasi FA* sebanyak 50 kali.

Setelah menentukan nilai parameter *FA* yang digunakan pada penelitian, maka didapatkan hasil evaluasi metode *SVR-FA* dengan menggunakan parameter pengujian model *SVR* terbaik yang telah diuji sebelumnya. Hasil evaluasi model *SVR-FA* dapat dilihat pada tabel X.

TABEL X. EVALUASI MODEL *SVR-FA*

RMSE	*MAE*	*MAPE%*
76.96	54.42	2.55%

Selain evaluasi akhir, pengujian *time series cross-validation* sebanyak 5 *fold* juga dilakukan pada model *SVR-FA* untuk mengukur tingkat konsistensi dan stabilitas model terhadap variasi pembagian data latih dan data validasi. Hasil pengujian cross-validation untuk model *SVR-FA* ddapat dilihat pada Tabel XI.

TABEL XI. PENGUJIAN CROSS VALIDATION SVR-FA

Fold	Data Latih	Data Validasi	*RMSE*	*MAE*	*MAPE%*
1	160	160	55.52	43.59	1.83%
2	320	160	76.46	56.85	2.61%
3	480	160	42.07	31.30	1.55%
4	640	160	27.70	20.39	1.10%
5	800	160	35.63	25.55	1.70%

Berdasarkan Tabel XI, terlihat bahwa tingkat kesalahan prediksi pada setiap *fold* yang dihasilkan oleh algoritma *SVR-FA* terjaga dengan baik tanpa mengalami lonjakan *error* yang drastis pada *fold* tertentu.

3.4 Perbandingan Hasil Evaluasi

Perbandingan hasil evaluasi dilakukan untuk mengetahui perbedaan kinerja performa antara model *SVR* dan model *hybrid SVR–FA* dalam memprediksi harga saham PT. Aneka Tambang Tbk. Perbandingan hasil evaluasi dapat dilihat pada Tabel XII.

Tabel XII. Perbandingan Hasil Evaluasi

Model	*RMSE*	*MAE*	*MAPE%*
SVR	93.17	63.30	2.77%
SVR-FA	76.96	54.42	2.55%

3.5 Pembahasan

Penelitian ini menggunakan model *hybrid Support Vector Regression – Firefly Algorithm* untuk memprediksi harga saham PT. Aneka Tambang Tbk. Data yang digunakan diperoleh dari situs *investing.com* dengan rentang waktu dari 11 September 2020 sampai 11 September 2025 dengan jumlah data sebanyak 1202 baris. Proses penelitian meliputi beberapa tahapan penting, mulai dari studi literatur, pengumpulan data, pre-*processing*, pemodelan *SVR* dan *SVR-FA*, serta evaluasi model. Model *SVR-FA* berhasil menurunkan kesalahan dengan *RMSE* 76.96, *MAE* 54.42, dan *MAPE* 2.55%, jauh lebih baik daripada *SVR* tanpa optimasi yang memiliki *RMSE* 93.17,

MAE 63.30, dan *MAPE* 2.77%. Keberhasilan ini membuktikan efektivitas *Firefly Algorithm* dalam mengatasi kelemahan *SVR*, yaitu sensitivitas terhadap parameter *hyperplane*. Temuan ini sejalan dengan penelitian sebelumnya yang menunjukkan bahwa optimasi *metaheuristik* pada *SVR* dapat menghasilkan akurasi lebih tinggi dibandingkan metode konvensional [25].

Peningkatan performa pada *SVR-FA* terjadi karena *FA* mengeksplorasi ruang parameter secara kontinu dan simultan, mengatasi kelemahan *SVR* standar yang terjebak pada pengujian manual rentang diskrit [11, 25]. Melalui daya tarik kunang-kunang dan pergerakan acak yang terukur *(alpha 0.05, gamma_fa 2.0, beta0 0.8)*, *FA* mampu menghindari *local minima* dan secara presisi menemukan koordinat *hyperplane* terbaik (kombinasi *C, epsilon, gamma*) untuk memodelkan data *nonlinier* saham ANTM [12, 22].

Peningkatan ini terbukti konsisten di seluruh pengujian data, yang dievaluasi dengan *time series cross-validation (5 fold)*. Berbeda dengan *SVR* standar yang sensitif dan rentan mengalami lonjakan *error* saat menerima data temporal baru, mekanisme *fitness (RMSE)* pada *SVR-FA* memastikan pemilihan parameter yang paling tangguh *(robust)* [25]. Hal ini dibuktikan dengan nilai standar deviasi lintas *fold* pada *SVR-FA* yang lebih rendah dan stabil, menegaskan kemampuannya dalam menurunkan kesalahan prediksi sekaligus menjaga kestabilan kinerja di berbagai skenario pembagian data.

4. KESIMPULAN DAN SARAN

4.1 Kesimpulan

Berdasarkan hasil penelitian, model *hybrid SVR-FA* terbukti meningkatkan akurasi prediksi harga saham PT Aneka Tambang Tbk dibandingkan *SVR* tanpa optimasi. Optimasi parameter menggunakan *Firefly Algorithm* secara konsisten menurunkan *RMSE*, *MAE*, dan *MAPE*, sementara penerapan *time series cross-validation* memberikan evaluasi kinerja yang lebih stabil dan objektif. Hasil ini menunjukkan bahwa *Firefly Algorithm* merupakan metode optimasi yang efektif untuk meningkatkan kinerja *SVR* pada prediksi harga saham berbasis data deret waktu.

4.2 Saran

Untuk penelitian selanjutnya, disarankan menerapkan model *hybrid SVR-FA* pada saham sektor lain atau indeks saham berbeda untuk menguji konsistensi kinerja. Penambahan variabel eksternal seperti indikator teknikal, data makroekonomi, atau sentimen pasar juga dapat meningkatkan akurasi prediksi. Selain itu, perbandingan *Firefly Algorithm* dengan metode optimasi *metaheuristik* lain, seperti *Particle Swarm Optimization* atau *Grey Wolf Optimizer*, dapat dilakukan untuk memperoleh pendekatan optimasi yang lebih optimal

5. UCAPAN TERIMAKASIH

Penulis mengucapkan terima kasih kepada semua pihak yang telah memberikan dukungan dan bantuan selama pelaksanaan penelitian ini. Penulis berharap karya ini dapat memberikan manfaat yang luas sekaligus menjadi referensi bagi studi-studi selanjutnya

Daftar Pustaka:

[1] S. D. Nabella, A. Munandar, and R. Tanjung, "Likuiditas, Solvabilitas, Aktivitas Dan Profitabilitas Terhadap Harga Saham Pada Perusahaan Sektor Tambangan Batu Bara Yang Terdaftar Di Bursa Efek Indonesia," *Meas. J. Akunt.*, vol. 16, no. 1, pp. 97–102, Jun. 2022, doi: 10.33373/mja.v16i1.4264.

[2] IDX Channel, "Saham ANTM Paling Diburu Asing Pekan Ini, Nilai Tembus Rp118 Miliar," IDX Channel. [Online]. Available: https://www.idxchannel.com/market-news/saham-antm-paling-diburu-asing-pekan-ini-nilai-tembus-rp118-miliar

[3] M. Darwish, E. E. Hassanien, and A. H. B. Eissa, "Stock Market Forecasting: From Traditional Predictive Models to Large Language Models," *Comput. Econ.*, vol. 67, no. 6, pp. 4553–4597, Jun. 2026, doi: 10.1007/s10614-025-11024-w.

[4] D. Muchlinski, "Machine learning and deep learning," in *Elgar Encyclopedia of Technology and Politics*, Edward Elgar Publishing, 2022, pp. 114–118. doi: 10.4337/9781800374263.machine.learning.deep.

[5] M. Bansal, A. Goyal, and A. Choudhary, "Stock Market Prediction with High Accuracy using Machine Learning Techniques," *Procedia Comput. Sci.*, vol. 215, pp. 247–265, 2022, doi: 10.1016/j.procs.2022.12.028.

[6] H. Oukhouya and K. El Himdi, "Comparing Machine Learning Methods—SVR, XGBoost, LSTM, and MLP— For Forecasting the Moroccan Stock Market," p. 39, 2023, doi: 10.3390/iocma2023-14409.

[7] D. S. Metwally, M. Ali, S. M. Alghamdi, and D. M. Khan, "A novel hybrid model to forecast the stock price based on CEEMDAN and support vector regression," *J. Radiat. Res. Appl. Sci.*, vol. 18, no. 2, p. 101385, 2025, doi: 10.1016/j.jrras.2025.101385.

[8] C. Comito and C. Pizzuti, "Artificial intelligence for forecasting and diagnosing COVID-19 pandemic: A focused review," *Artif. Intell. Med.*, vol. 128, no. March, p. 102286, 2022, doi: 10.1016/j.artmed.2022.102286.

[9] R. Guido, S. Ferrisi, D. Lofaro, and D. Conforti, "An Overview on the Advancements of Support Vector Machine Models in Healthcare Applications: A Review," *Inf.*, vol. 15, no. 4, 2024, doi: 10.3390/info15040235.

[10] I. F. Riziq and A. R. Dzikrillah, "Impelementasi Algoritma LSTM Dan SVR Untuk Prediksi Harga Bitcoin Menggunakan Data Yahoo Finance," *Metr. J.*, vol. 9, no. 2, pp. 284–291, 2025, doi: 10.47002/metik.v9i2.1077.

[11] M. Saber, A. A. Abdelhamid, and A. Ibrahim, "Metaheuristic Optimization Review: Algorithms and Applications," *J. Artif. Intell. Metaheuristics*, vol. 3, no. 1, pp. 21–30, 2023, doi: 10.54216/JAIM.030102.

[12] W. Li, W. Li, and Y. Huang, "Enhancing Firefly Algorithm with Dual-Population Topology Coevolution," *Mathematics*, vol. 10, no. 9, p. 1564, May 2022, doi: 10.3390/math10091564.

[13] M. Achite, S. Samadianfard, N. Elshaboury, K. A. Toubal, E. M. Abdelkader, and M. Sharafi, "A combined support vector regression with a firefly algorithm for prediction of energy consumption in wastewater treatment plants," *Water Sci. Technol.*, vol. 90, no. 10, pp. 2747–2763, Nov. 2024, doi: 10.2166/wst.2024.375.

[14] Siti Muawwanah, Taghfirul Azhima Yoga Siswa, and Wawan Joko Pranoto, "Model Optimasi SVM-GSBE dalam Menangani High Dimensional Data Stunting Kota Samarinda," *J. Teknol. Sist. Inf. dan Apl.*, vol. 7, no. 3, pp. 1246–1258, 2024, doi: 10.32493/jtsi.v7i3.41545.

[15] E. N. Weinstein and J. W. Miller, "Bayesian Data Selection," *J. Mach. Learn. Res.*, vol. 24, pp. 1–72, 2023.

[16] C. Yulian *et al.*, "Analisis Implementasi Teknologi Blockchain dalam Meningkatkan Transparansi dan Kepercayaan di Sektor Keuangan di Indonesia," *J. Sist. Inf. dan Ilmu Komput.*, no. 4, pp. 9–31, 2024.

[17] O. Surakhi *et al.*, "Time-lag selection for time-series forecasting using neural network and heuristic algorithm," *Electron.*, vol. 10, no. 20, pp. 1–22, 2021, doi: 10.3390/electronics10202518.

[18] M. Sholeh, D. Andayati, and R. Y. Rachmawati, "Data Mining Model Klasifikasi Menggunakan Algoritma K-Nearest Neighbor Dengan Normalisasi Untuk Prediksi Penyakit Diabetes," *TeIKa*, vol. 12, no. 02, pp. 77–87, 2022, doi: 10.36342/teika.v12i02.2911.

[19] Z. Annisa, N. Azizah, I. Cholissodin, and L. Muflikhah, "Prediksi Hasil Panen Tanaman Biofarmaka di Indonesia dengan Menggunakan Metode Extreme Learning Machine," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 6, no. 11, pp. 5331–5338, 2022.

[20] A. M. Priyatno, L. S. Tanjung, W. F. Ramadhan, P. Cholidhazia, P. Z. Jati, and F. I. Firmananda, "Comparison Random Forest Regression and Linear Regression For Forecasting BBCA Stock Price," *J. Tek. Ind. Terintegrasi*, vol. 6, no. 3, pp. 718–732, Jul. 2023, doi: 10.31004/jutin.v6i3.16933.

[21] S. Saadah, F. Z. Z, and H. H. Z, "Support Vector Regression (SVR) Dalam Memprediksi Harga Minyak Kelapa Sawit di Indonesia dan Nilai Tukar Mata Uang EUR/USD," *J. Comput. Sci. Informatics Eng.*, vol. 5, no. 1, pp. 85–92, 2021, doi: 10.29303/jcosine.v5i1.403.

[22] Y. Li, Y. Zhao, Y. Shang, and J. Liu, "An improved firefly algorithm with dynamic self-adaptive adjustment," *PLoS One*, vol. 16, no. 10 October 2021, pp. 1–24, 2021,

doi: 10.1371/journal.pone.0255951.

[23] C. He, Z. Li, C. Zheng, Z. Zhang, and L. Zhang, "A Hybrid Model Based on a Dual-Attention Mechanism for the Prediction of Remaining Useful Life of Aircraft Engines," pp. 1–25, 2025.

[24] A. Hajjah and Y. N. Marlim, "Analisis Error Terhadap Peramalan Data Penjualan," *Techno.Com*, vol. 20, no. 1, pp. 1–9, 2021, doi: 10.33633/tc.v20i1.4054.

[25] A. Pradjaningsih and L. H. R. Romadhoni, "Firefly Algorithm-Optimized SVR Framework for Accurate Stock Price Forecasting," *Stat. Optim. Inf. Comput.*, vol. 14, no. 3, pp. 1403–1418, 2025, doi: 10.19139/soic-2310-5070-2387.