

Optimasi Parameter Support Vector Regression (SVR) dengan Grid Search Optimization (GSO) dan Seleksi Fitur dalam Prediksi Harga Saham

Nia Andriani Laila

*a Universitas Ahmad Dahlan, Yogyakarta, Indonesia;

¹ nia2000015002@webmail.uad.ac.id

*Correspondent Author

Received:

Revised:

Accepted:

KATAKUNCI

Prediksi
Corelation Based Feature
Selection (CFS)
Grid Search Optimization (GSO)
Support Vector Regression
(SVR)

KEYWORDS

Prediction
Corelation Based Feature
Selection (CFS)
Grid Search Optimization
(GSO)
Support Vector Regression
(SVR)

ABSTRAK

Prediksi adalah suatu metode yang dilakukan untuk mendapatkan gambaran atau memperkirakan sesuatu yang terjadi di masa yang akan datang dengan menggunakan informasi atau data yang ada pada masa lampau dan masa kini. Support Vector Regression (SVR) merupakan salah satu metode yang digunakan untuk melakukan prediksi. Penelitian ini dilakukan dengan tujuan untuk meningkatkan performa dari SVR dalam memprediksi harga saham pada periode 1 februari 2021 hingga 23 februari 2023. Oleh karena itu pada penelitian ini menggunakan seleksi fitur Corelation Based Feature Selection (CFS) dan metode optimasi Grid Search Optimization (GSO) yang diimplementasikan pada SVR. Hasil penelitian menunjukkan bahwa prediksi model SVR diperoleh nilai RMSE sebesar 0.00703 pada data training dan 0.00611 pada data testing. Sedangkan pada model SVR dengan algoritma GSO diperoleh nilai RMSE sebesar 0.00429 pada data training dan 0.00367 pada data testing. Berdasarkan nilai RMSE yang diperoleh menunjukkan adanya peningkatan performa pada SVR dengan penurunan nilai RMSE sebesar 0.00274 pada data training dan 0.00244 pada data testing.

ABSTRACT

Prediction is a method used to get a picture of or estimate something that will happen in the future using information or data that exists in the past and present. Support Vector Regression (SVR) is one of the methods used to make predictions. This research was conducted with the aim of improving the performance of SVR in predicting stock prices for the period February 1, 2021, to February 23, 2023. Therefore, this study uses the correlation-based feature selection (CFS) and grid search optimization (GSO) optimization methods implemented in SVR. The results showed that the prediction of the SVR model obtained an RMSE value of 0.00703 on training data and 0.00611 on testing data. While the SVR model with the GSO algorithm obtained an RMSE value of 0.00429 in the testing data and 0.00367 in the testing data, based on the RMSE value obtained, it shows an increase in performance in SVR with a decrease in the RMSE value of 0.00274 in the training data and 0.00244 in the testing data.

This is an open-access article under the [CC-BY-SA](#) license.



Pendahuluan

Prediksi merupakan metode untuk memperkirakan kejadian di masa depan berdasarkan data masa lalu dan masa kini. Meskipun tidak menjamin hasil yang akurat, prediksi bertujuan memberikan estimasi yang mendekati kondisi nyata. Oleh karena itu, diperlukan teknik prediksi yang tepat agar hasilnya optimal. Dalam hal ini, data mining dan machine learning memegang peran penting, terutama dalam kasus data runtun waktu (time series) [1].

Beberapa metode machine learning yang umum digunakan dalam prediksi antara lain Support Vector Regression (SVR), ARIMA, dan Random Forest (RF). SVR telah menunjukkan performa yang baik dalam berbagai penelitian. Misalnya, [2] menunjukkan bahwa SVR lebih baik dibandingkan RF dalam memprediksi tegangan output SOFC berdasarkan nilai MAPE. Penelitian lainnya [3] juga menunjukkan bahwa SVR mengungguli ARIMA berdasarkan nilai RMSE pada tiga dataset berbeda.

Namun, kinerja SVR sangat bergantung pada pemilihan parameter yang optimal [4][5]. Grid Search Optimization (GSO) merupakan salah satu metode pencarian parameter terbaik yang banyak digunakan [6]. Penelitian [7] menggunakan GSO pada model deep learning (LSTM dan CNN) untuk memprediksi beban listrik dan memperoleh nilai MAPE yang baik. Sementara itu, [8] menerapkan GSO pada model KNN untuk prediksi emisi karbon dan mendapatkan nilai MSE yang cukup rendah. Penelitian [9] juga menggabungkan SVR dan Improved Grid Search (IGS) untuk memprediksi pengembangan alat analisa pertanian, namun menyarankan penggunaan GSO untuk hasil yang lebih baik.

Selain optimasi parameter, seleksi fitur juga merupakan langkah penting dalam membangun model prediksi yang kuat. Seleksi fitur membantu memilih fitur yang relevan dan membuang yang tidak relevan [10]. Dalam penelitian [11], model SVR yang menggunakan seleksi fitur menunjukkan hasil yang lebih baik berdasarkan nilai MSPE dibandingkan tanpa seleksi fitur. Salah satu metode seleksi fitur yang efektif adalah Correlation Based Feature Selection (CFS), yang menilai korelasi antar fitur. Penelitian [12] membandingkan CFS dan PCA dalam deteksi penyakit hati dengan metode MLP, dan hasilnya menunjukkan bahwa CFS menghasilkan akurasi yang lebih tinggi.

Berdasarkan hasil kajian tersebut, penelitian ini bertujuan menggabungkan metode Support Vector Regression (SVR) dengan optimasi parameter menggunakan Grid Search Optimization (GSO) serta seleksi fitur menggunakan Correlation Based Feature Selection (CFS) dalam memprediksi harga saham.

Metode

Penelitian ini merupakan penelitian kuantitatif dengan pendekatan eksperimen yang bertujuan untuk memprediksi harga saham menggunakan algoritma Support Vector Regression (SVR) yang dioptimasi dengan Grid Search Optimization (GSO) dan seleksi fitur Correlation Based Feature Selection (CFS). Data yang digunakan adalah data historis harga saham yang diperoleh dari sumber terbuka atau publik, kemudian dikumpulkan melalui metode dokumentasi. Data yang telah diperoleh selanjutnya dinormalisasi untuk menyamakan skala antar fitur, lalu dilakukan seleksi fitur menggunakan metode CFS guna memilih fitur yang relevan terhadap target prediksi. Model SVR dibangun dengan tahapan pemilihan kernel yang sesuai, dilanjutkan dengan proses optimasi parameter menggunakan GSO guna memperoleh kombinasi parameter terbaik. Setelah model dibangun, dilakukan pengujian terhadap data uji dan evaluasi kinerja model dilakukan menggunakan metrik Root Mean Square Error (RMSE). Simpulan dari penelitian ini ditarik berdasarkan hasil evaluasi model terhadap data uji yang menunjukkan seberapa akurat dan efektif pendekatan SVR dengan GSO dan CFS dalam melakukan prediksi harga saham.

Hasil dan Pembahasan

1. Pengolahan Data

Penelitian ini menggunakan dataset dari situs web yahoo!finance dengan alamat website <https://finance.yahoo.com/quote/BRIS.JK/>. Dataset tersebut berisi tentang harga harian saham PT BSI Tbk pada 5 hari kerja yang diambil sejak tanggal 1 Februari 2021 hingga 23 Februari 2023 dengan data berjumlah sebanyak 511 data. Dataset tersebut terdiri dari 7 fitur yaitu Date (tanggal), Open (harga pembuka), High (harga tertinggi), Low (harga terendah), Close (harga penutupan), Adj Close (harga penutupan yang disesuaikan), dan Volume (jumlah saham yang diperdagangkan). Selanjutnya dilakukan proses preprocessing yang diawali dengan analisis statistik deskriptif yang disajikan pada tabel 1.

Berdasarkan tabel 1 diperoleh bahwa nilai Count atau jumlah data yang valid ada sebanyak 511 data. Oleh karena itu semua data pada data harga saham dapat diproses dan tidak ada data yang hilang atau tidak mengandung NaN. Proses selanjutnya akan dilakukan perhitungan

normalisasi dan hasilnya disajikan pada tabel 2 sebagai berikut :

Tabel 1. Hasil Statistik Deskriptif

<i>Variabel</i>	<i>Count</i>	<i>Minimum</i>	<i>Maximum</i>	<i>Mean</i>	<i>Std. Deviation</i>
Open	511	1100.000	2906.420	1777.704	432.756
High	511	1140.000	2994.198	1814.354	448.612
Low	511	1095.000	2808.889	1742.444	415.908
Close	511	1100.000	2886.914	1772.182	430.134
Adj Close	511	1084.937	2811.485	1732.135	413.512
Volume	511	2.300604e+06	8.082770e+08	4.046846e+07	6.680163e+07

Tabel 2. Hasil Normalisasi Data

<i>Open</i>	<i>High</i>	<i>Low</i>	<i>Close</i>	<i>Adj Close</i>	<i>Volume</i>
0.713846	0.942140	0.652872	0.912671	0.911979	1.000000
0.919013	0.894800	0.846353	0.808968	0.807455	0.388090
0.800232	0.863240	0.783757	0.885381	0.884473	0.525672
0.886618	0.879020	0.852044	0.847174	0.845964	0.307529
⋮	⋮	⋮	⋮	⋮	⋮
0.254647	0.264265	0.224635	0.293803	0.299911	0.217385

2. Implementasi Algoritma Correlation Based Feature Selection (CFS)

Setelah Normalisasi data langkah selanjutnya dari *preprocessing* adalah seleksi fitur. Seleksi fitur yang digunakan dalam penelitian ini adalah *Correlation Based Feature Selection*(CFS). Langkah yang digunakan yaitu menghitung korelasi pearson, menghitung merit, dan menentukan fitur optimal. Setelah diperoleh nilai korelasi pearson, kemudian menggabungkan korelasi fitur dengan fitur target (relevansi) dan mengurangnya dengan penalti redundansi.

Tahapan selanjutnya yaitu memilih fitur terbaik yang diinginkan berdasarkan nilai merit yang telah diurutkan. Pada penelitian ini, akan diambil ambang batas untuk nilai merit yaitu 0.5. Sehingga fitur-fitur yang diambil adalah fitur yang memiliki nilai merit . Oleh karena itu, diperoleh 4 fitur terbaik yang telah direduksi. Dengan demikian diperoleh dataset baru dengan fitur open, high, low, adj close yang merupakan fitur yang berkorelasi dengan fitur target dan juga tidak berkorelasi dengan fitur-fitur lainnya. Dataset terbaru disajikan pada tabel 3 berikut:

Tabel 3. Dataset Setelah Seleksi Fitur

<i>Open</i>	<i>High</i>	<i>Low</i>	<i>Adj Close</i>	<i>Close</i>
0.713846	0.942140	0.652872	0.911979	0.912671
0.919013	0.894800	0.846353	0.807455	0.808968
0.800232	0.863240	0.783757	0.884473	0.885381
0.886618	0.879020	0.852044	0.845964	0.847174
⋮	⋮	⋮	⋮	⋮
0.254647	0.264265	0.224635	0.299911	0.293803

3. Implementasi Support Vector Regression (SVR)

Pada metode SVR memiliki beberapa tahapan proses yaitu, melakukan split data atau membagi data menjadi data training dan data testing, menentukan kernel, menentukan nilai C , epsilon dan parameter kernel, dan hasil prediksi dari data training dan data testing.

a. Pemisahan Data Training dan Data Testing

Pada penelitian ini dataset dibagi menjadi 80% data training dan 20% data testing. Hasil pembagian dataset tersebut diperoleh 408 data training dan 103 data testing.

b. Menentukan Kernel

Kernel RBF berfungsi untuk menentukan hubungan antara variabel independen dan dependen. Kernel ini juga digunakan untuk menghitung jarak antara support vector dan titik data lainnya dalam ruang fitur yang telah diperluas. Selanjutnya, nilai yang dihasilkan dari kernel RBF digunakan dalam formulasi matematis SVR untuk menghitung prediksi. Fungsi kernel RBF ini dapat dihitung menggunakan persamaan 1 :

$$k(x_i, x_j) = \exp(-\gamma(\|x_i - x_j\|)^2) \quad (1)$$

Setelah perhitungan kernel RBF dilakukan dengan keseluruhan data training, akan diperoleh nilai kernel dengan matriks sebagai berikut :

$$K_{408 \times 408} = \begin{bmatrix} 1 & 0.999907 & 0.999968 & \dots & 0.998380 \\ 0.999907 & 1 & 0.999975 & \dots & 0.998108 \\ 0.999968 & 0.999975 & 1 & \dots & 0.998291 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0.998380 & 0.998108 & 0.998291 & \dots & 1 \end{bmatrix}$$

c. Membentuk Model SVR

Dalam membentuk model SVR, diperlukan penentuan beberapa parameter penting. Di antaranya adalah parameter C , epsilon (ϵ), dan gamma (γ). Nilai parameter tersebut akan disajikan pada tabel 4.10 berikut.

Tabel 4. Nilai Parameter C , Epsilon, dan Gamma

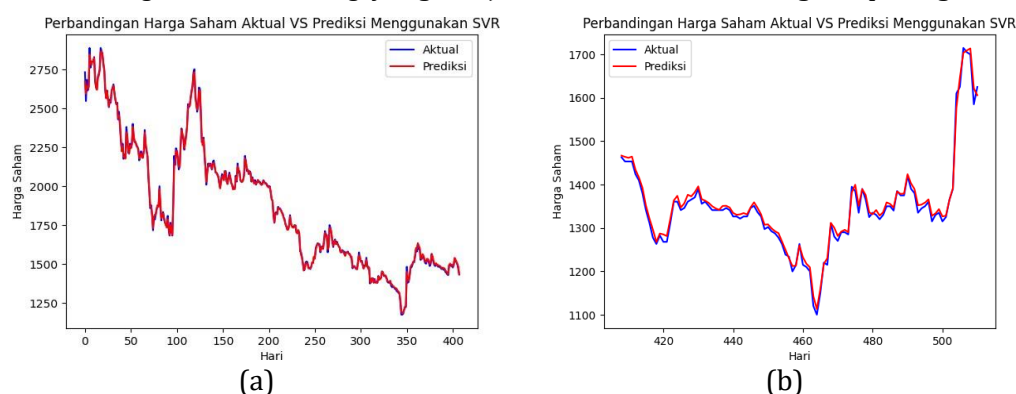
C	(ϵ)	(γ)
100	0.01	0.001

Dari parameter tersebut digunakan untuk mencari lagrange multiplier sehingga diperoleh nilai bobot dan bias yang disajikan pada tabel 5 berikut:

Tabel 5. Nilai Bobot dan Bias

<i>Parameter</i>	<i>Nilai Parameter</i>
w_1	-14.48895428
w_2	133.03864044
w_3	90.12684854
w_4	288.24023404
b	0.59014257

Setelah memperoleh model dan parameter yang digunakan, selanjutnya mencari nilai prediksi pada *data training* dan *data testing* yang disajikan dalam bentuk diagram pada gambar 1:



Gambar 1. Hasil Prediksi Harga Saham

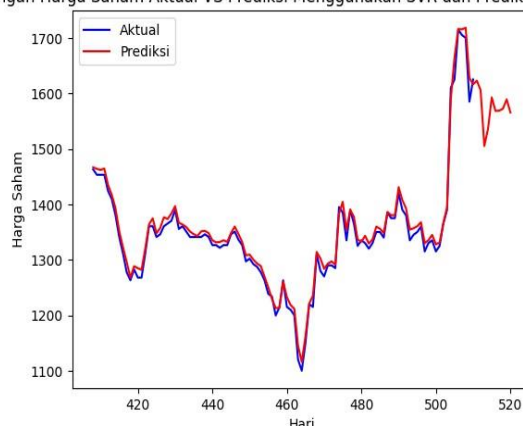
Berdasarkan gambar 1, grafik dengan garis merah menunjukkan pola hasil prediksi harga penutupan saham, sementara garis biru menunjukkan pola data harga penutupan saham aktual. Garis biru digunakan untuk memperjelas keakuratan prediksi dengan melihat perpotongan antara kedua data tersebut. Berdasarkan grafik tersebut, dapat disimpulkan bahwa model SVR dengan parameter yang diuji memiliki kemampuan prediksi yang cukup baik. Hal ini ditunjukkan oleh perpotongan grafik antara data aktual dan prediksi yang hampir berhimpit di beberapa titik. Setelah memperoleh nilai prediksi harga penutupan saham, dilakukan prediksi untuk 10 hari ke depan. Tujuannya adalah untuk memastikan apakah prediksi yang telah dilakukan bekerja dengan baik dan akurat. Hasil dari prediksi tersebut disajikan pada tabel 6 dan pada gambar 2 berikut ini.

Tabel 6. Hasil Prediksi 10 Hari Ke depan

<i>Data Aktual</i>	<i>Data Prediksi</i>
1630	1607.70615334
1520	1607.84980875
1520	1487.18192081
1595	1507.97806058
1565	1579.7316951

1565	1554.78880069
1565	1563.62330047
1595	1550.51805548
1555	1583.46789293
1595	1545.63802328

Perbandingan Harga Saham Aktual VS Prediksi Menggunakan SVR dan Prediksi 10 Hari ke Depan



Gambar 2. Grafik perbandingan Data aktual dan Prediksi

Berdasarkan Gambar 2 di atas, prediksi harga penutupan saham untuk 10 hari ke depan menunjukkan tren penurunan. Hal ini cukup sesuai dengan harga aktualnya. Dengan demikian, dapat disimpulkan bahwa model SVR ini cukup baik, karena hasil prediksinya cukup sesuai dengan data aktual yang mengalami penurunan setelah 10 hari ke depan.

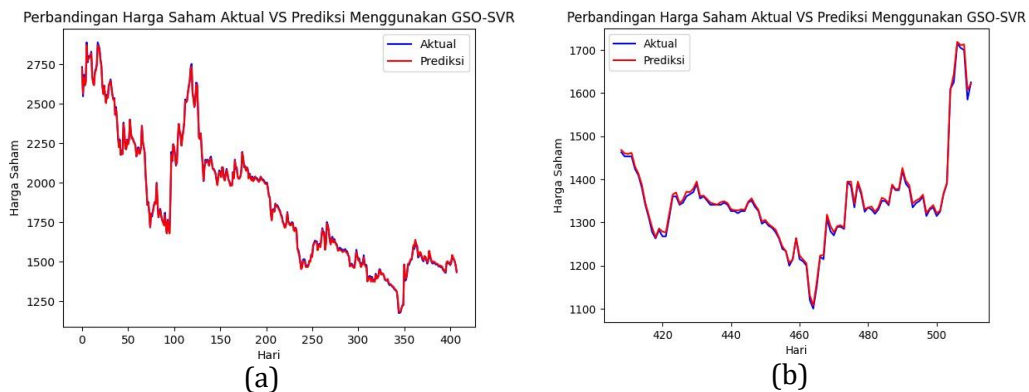
4. SVR dengan Grid Search Optimization (GSO)

Algoritma GSO digunakan untuk memperbarui dan mengoptimalkan parameter-parameter seperti ϵ , dan γ selama pelatihan model SVR. Dengan menggunakan metode grid search optimization (GSO), kita dapat meningkatkan performa model SVR untuk prediksi data harga saham dengan menemukan hyperparameter yang paling optimal. Setelah parameter optimal ditemukan, model diuji dan dievaluasi menggunakan data uji untuk memastikan akurasi. Berikut adalah hasil parameter terbaik dari perhitungan GSO.

Table 7. Nilai Parameter Menggunakan GSO

C	(ϵ)	(γ)
1000	0.01	0.01

Setelah diperoleh nilai parameter terbaik, langkah selanjutnya yaitu melakukan prediksi harga saham menggunakan nilai parameter tersebut. Hasil prediksi direpresentasikan pada diagram berikut.



Gambar 3. Hasil Prediksi Harga Saham

Berdasarkan Gambar 3, grafik dengan garis merah menunjukkan pola hasil prediksi harga saham, sementara garis biru menggambarkan pola data harga penutupan saham aktual. Garis biru digunakan untuk memperjelas keakuratan prediksi melalui perpotongan antara kedua data tersebut. Berdasarkan grafik tersebut, terlihat bahwa model SVR-GSO dengan parameter yang diuji memiliki kemampuan prediksi yang lebih baik dibandingkan dengan model SVR biasa. Setelah memperoleh nilai prediksi harga penutupan saham, prediksi untuk 10 hari ke depan dilakukan untuk memastikan apakah prediksi tersebut bekerja dengan baik dan akurat. Hasil dari prediksi tersebut disajikan pada gambar 4 dan tabel 8 berikut ini.

Berdasarkan Gambar 6 dan Tabel 8, prediksi harga penutupan saham untuk 10 hari ke depan menunjukkan tren penurunan. Prediksi ini lebih sesuai 53 dengan harga aktual dibandingkan dengan metode SVR. Dengan demikian, dapat disimpulkan bahwa model SVR dengan algoritma GSO ini cukup baik, karena hasil prediksinya cukup sesuai dengan data aktual yang mengalami penurunan setelah 10 hari ke depan.



Gambar 6. Grafik perbandingan Data aktual dan Prediksi

Tabel 8. Hasil Prediksi 10 Hari Ke depan

<i>Data Aktual</i>	<i>Data Prediksi</i>
1630	1631.18813688
1520	1548.6195832
1520	1516.40976727
1595	1587.1158371
1565	1582.49990016
1565	1573.32325637
1565	1568.00359864
1595	1595.87850416
1555	1570.41884084
1595	1594.74298416

5. Evaluasi Model

Hasil dan Pembahasan dapat disajikan dalam subbab. Membahas secara jelas pokok bahasan sesuai dengan masalah, tujuan penelitian, dan teori yang digunakan.

Evaluasi model dilakukan dengan menghitung nilai RMSE dari model SVR dan model SVR dengan GSO yang disajikan pada tabel 9.

Table 9. Hasil RMSE

<i>Metode</i>	<i>RMSE</i>	
	<i>Data Training</i>	<i>Data Testing</i>
SVR	0.00703	0.00611
SVR-GSO	0.00429	0.00367

Berdasarkan hasil RMSE Tabel 9, terlihat bahwa metode SVR dan SVR dengan GSO menghasilkan nilai RMSE yang berbeda. Metode SVR memiliki nilai RMSE sebesar 0.00703 untuk data pelatihan dan 0.00611 untuk data pengujian. Sementara itu, metode SVR dengan GSO memiliki nilai RMSE sebesar 0.00429 untuk data pelatihan dan 0.00367 untuk data pengujian. Hal ini menunjukkan adanya penurunan nilai RMSE sebesar 0.00274 untuk data pelatihan dan 0.00244 untuk data pengujian. Dengan demikian, dapat disimpulkan bahwa metode SVR dengan GSO mampu meningkatkan performa prediksi metode SVR.

Simpulan

Penerapan metode seleksi fitur menggunakan Correlation Based Feature Selection (CFS) berhasil mereduksi lima variabel independen menjadi empat variabel (Open, High, Low, dan Adj Close) yang memiliki korelasi tinggi terhadap target namun rendah terhadap sesama fitur. Model prediksi harga saham menggunakan Support Vector Regression (SVR) yang dioptimasi dengan algoritma Grid Search Optimization (GSO) menunjukkan peningkatan performa signifikan dibandingkan model SVR tanpa optimasi. Nilai Root Mean Square Error (RMSE) pada

model SVR tanpa GSO adalah 0.00703 untuk data pelatihan dan 0.00611 untuk data pengujian, sedangkan pada model SVR dengan GSO diperoleh RMSE sebesar 0.00429 untuk data pelatihan dan 0.00367 untuk data pengujian. Penurunan nilai RMSE sebesar 0.00274 pada data pelatihan dan 0.00244 pada data pengujian menunjukkan bahwa kombinasi CFS dan GSO mampu meningkatkan akurasi model SVR dalam memprediksi harga saham pada periode 1 Februari 2021 hingga 23 Februari 2023, sehingga metode ini layak dijadikan alternatif dalam sistem prediksi saham.

Daftar Pustaka

- [1] Y. Kusumawati, K. Widyatmoko, and C. Irawan, "Gold Price Prediction Using Support Vector Regression," *J. Appl. Intell. Syst.*, vol. 7, no. 1, pp. 89–102, 2022, doi: 10.33633/jais.v7i1.6124.
- [2] F. C. İskenderoğlu, M. K. Baltacıoglu, M. H. Demir, G. Bidini, A. Baldinelli, and L. Barelli, "Comparison of support vector regression and random forest algorithms for estimating the SOFC output voltage by considering hydrogen flow rates," *Int. J. Hydrog. Energy*, vol. 45, no. 60, pp. 35023–35038, 2020, doi: 10.1016/j.ijhydene.2020.07.265.
- [3] Y. S. Amirkhalili, A. Aghsami, and F. Jolai, "Comparison of Time Series ARIMA Model and Support Vector Regression," *Int. J. Hybrid Inf. Technol.*, vol. 13, no. 1, pp. 7–18, 2020, doi: 10.21742/ijhit.2020.13.1.02.
- [4] C. Peng, Z. Che, T. W. Liao, and Z. Zhang, "Prediction using multi-objective slime mould algorithm optimized support vector regression model," *Appl. Soft Comput.*, vol. 145, p. 110580, 2023, doi: 10.1016/j.asoc.2023.110580.
- [5] Sukamto, Hadiyanto, and Kurnianingsih, "KNN Optimization Using Grid Search Algorithm for Preeclampsia Imbalance Class," *E3S Web Conf.*, vol. 448, 2023, doi: 10.1051/e3sconf/202344802057.
- [6] S. Shinde, "Liver Disease Prediction Based on Grid Search and Random Forest Classification," *Int. J. Eng. Appl. Sci. Technol.*, vol. 7, no. 1, pp. 136–140, 2022, doi: 10.33564/ijeast.2022.v07i01.020.
- [7] T. T. Ngoc, L. Van Dai, and L. B. Minh, "Effects of Data Standardization on Hyperparameter Optimization with the Grid Search Algorithm Based on Deep Learning: A Case Study of Electric Load Forecasting," *Adv. Technol. Innov.*, vol. 7, no. 4, pp. 258–269, 2022, doi: 10.46604/aiti.2022.9227.
- [8] Y. Qi, H. Liu, and J. Zhao, "Prediction model and demonstration of regional agricultural carbon emissions based on Isomap-ACO-ET: a case study of Guangdong Province, China," *Sci. Rep.*, vol. 13, no. 1, pp. 1–11, 2023, doi: 10.1038/s41598-023-39996-5.
- [9] Y. Sun, S. Ding, and Z. Z. & W. Jia, "An improved grid search algorithm to optimize SVR for prediction," vol. 25, pp. 5633–5644, 2021.
- [10] A. N. Puteri, A. Arizal, and A. D. Achmad, "Feature Selection Correlation-Based pada Prediksi Nasabah Bank Telemarketing untuk Deposito," *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 20, no. 2, pp. 335–342, 2021, doi: 10.30812/matrik.v20i2.1183.
- [11] H. Fröhlich, O. Chapelle, and B. Schölkopf, "Feature Selection for Support Vector Machines using Genetic Algorithms," *Int. J. AI Tools Spec. Issue Sel. Pap. from 15th IEEE Int. Conf. Tools with AI 2003*, vol. 13, no. 4, pp. 791–800, 2004, doi: 10.1142/S0218213004001818.
- [12] A. M. Kuruvilla and N. . Balaji, "Heart disease prediction system using Correlation Based Feature Selection with Multilayer Perceptron approach," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1085, no. 1, p. 012028, 2021, doi: 10.1088/1757-899x/1085/1/012028.