

IMPLEMENTASI *GRIDSEARCHCV* PADA *SUPPORT VECTOR REGRESSION* (SVR) UNTUK PERAMALAN HARGA SAHAM (Studi Kasus: Harga Saham PT Anabatic Technologies Tbk)

Aanisah Waliy Ishlah^{1*}, Sudarno², Puspita Kartikasari³

^{1,2,3}Departemen Statistika, Fakultas Sains dan Matematika, Universitas Diponegoro

*e-mail: waliyaanisah@gmail.com

DOI: 10.14710/j.gauss.12.2.276-286

Article Info:

Received: 2022-11-16

Accepted: 2023-04-20

Available Online: 2023-07-28

Keywords:

Stock; Support Vector Regression; GridSearchCV; Kernel; Prediction

Abstract: Stock is a sign of the capital participation of a person or authority in a company (PT). PT Anabatic Technologies Tbk (ATIC) is one of the service providers and IT consultants that is included in the technology sector, which is a new sector in the IDX-IC classification. ATIC stock trading was temporarily suspended due to a significant increase in cumulative prices. This indicates that stock prices tend to be volatile and non-linear. The Support Vector Regression (SVR) method can be used to predict stock prices. SVR is able to solve non-linear data problems by using kernel functions so it can overcome overfitting problems and will give good performance. The SVR problem is difficult to determine the optimal hyperparameters, so this research uses grid search cross validation (GridSearchCV). In this research, ATIC's daily closing price data was used with 1007 training data and 252 testing data. The results show that the best model is SVR with a linear kernel and the hyperparameters used are Cost (C) = 1 and epsilon (ϵ) = 0,11. The linear kernel SVR model produces a MSE of 0,001237173; SMAPE of 0,1167301; and $R^2 = 0,9206643$.

1. PENDAHULUAN

Bursa Efek Indonesia (BEI) telah menerapkan klasifikasi sektor dan industri perusahaan baru, yaitu *Indonesia Stock Exchange Industrial Classification* (IDX-IC) sejak 25 Januari 2021 dimana terdapat 12 sektor, salah satunya adalah sektor teknologi (Bursa Efek Indonesia, 2018). Data pada BEI mengungkapkan bahwa indeks sektor teknologi berhasil meningkat 707,56% *year to date* (ytd) dan mengalami pertumbuhan tertinggi sepanjang tahun 2021. PT Anabatic Technologies Tbk adalah salah satu perusahaan dari sektor teknologi yang menyediakan jasa dan konsultan TI dan memiliki kode saham ATIC. Penjualan saham milik ATIC dihentikan sementara pada Senin, 21 Juni 2021 dikarenakan adanya peningkatan harga kumulatif yang signifikan. Harga sahamnya meningkat hingga 143% dalam empat hari, sehingga bursa melaporkan bahwa saham ATIC tergolong ke dalam *Unusual Market Activity* (UMA) (Intan, 2021). Hal tersebut menandakan bahwa harga saham cenderung fluktuatif dan non linier. Salah satu metode prediksi yang umum digunakan adalah metode regresi linier, namun metode tersebut identik dengan beberapa asumsi yang harus dipenuhi. *Support Vector Machine* (SVM) dapat digunakan untuk menangani permasalahan asumsi pada regresi linier. SVM yang digunakan pada kasus regresi disebut *Support Vector Regression* (SVR) (Saputra, *et al.*, 2019).

SVR adalah penggunaan dari *Support Vector Machine* (SVM) dalam kasus regresi yang dapat menangani masalah data non linier menggunakan fungsi kernel, sehingga dapat menangani permasalahan *overfitting* dan akan menunjukkan performansi yang bagus (Santosa, 2007). Permasalahan yang ditemukan pada SVR adalah sulit untuk menetapkan *hyperparameter* optimalnya untuk membangun model terbaik. Penetapan *hyperparameter*

yang optimal dapat dilakukan dengan menerapkan *grid search* yang dipadukan dengan *cross validation* (Saputra, *et al.*, 2019). Tingkat kesalahan prediksi, akurasi, dan kesesuaian pada hasil prediksi dapat dihitung menggunakan nilai MSE, SMAPE dan R^2 (Koefisien Determinasi).

2. TINJAUAN PUSTAKA

PT Anabatic Technologies Tbk adalah salah satu perusahaan yang terdapat pada sektor teknologi. Sektor teknologi merupakan salah satu sektor yang terdapat pada pengklasifikasian saham terbaru bernama IDX-IC. Emiten tersebut tercatat di BEI pada tanggal 08 Juli 2015 dengan kode saham ATIC.

Makridakis, *et al.* (1992) mengemukakan bahwa peramalan masa depan didasarkan terhadap nilai masa lalu dari variabel pada *time series*. Memperoleh pola dalam deret data masa lalu dan mengekstrapolasi pola tersebut ke masa depan merupakan tujuan dari peramalan *time series*.

Data mentah akan melewati *preprocessing* sebelum diproses dengan algoritma *machine learning*. *Preprocessing* data meliputi *data cleaning*, *data transformation*, dan *data reduction*. Proses *data cleaning* akan menghilangkan data yang terdeteksi terdapat *missing value* dan *outlier* (Antony, *et al.*, 2021). Proses *data transformation* meliputi normalisasi data dengan normalisasi *min-max* yang mengubah data menjadi rentang lebih kecil, seperti $[-1, 1]$ atau $[0, 1]$. Normalisasi *min-max* akan menjaga hubungan diantara nilai data asli karena melakukan transformasi linier pada data asli. Rumus normalisasi *min-max* seperti pada persamaan (1) (Han, *et al.*, 2012).

$$x'_i = \frac{x_i - x_{min}}{x_{max} - x_{min}} (maxR - minR) + minR \quad (1)$$

Dengan x'_i adalah data hasil normalisasi pada baris data ke- i , x_i adalah data aktual pada baris data ke- i , x_{min} adalah nilai minimum data aktual, x_{max} adalah nilai maksimum data aktual, $minR$ adalah rentang nilai terkecil yang baru, dan $maxR$ adalah rentang nilai terbesar yang baru.

Menurut Prabowo, *et al.* (2020), model statistik saat ini tidak hanya mencakup model linier tetapi juga model non linier. Penelitian ini menggunakan uji *Ramsey* dalam mendeteksi non linier yang didasarkan pada uji galat spesifikasi. Misalkan terdapat 2 variabel, yaitu variabel prediktor (X) dan variabel respon (Y). Model linier seperti pada persamaan (2).

$$Y_i = \beta_1 + \beta_2 X_i + u_i \quad (2)$$

Jika \hat{Y}_i dan \hat{u}_i diplotkan dan masih terlihat pola, maka model tidak sesuai. Hal ini berarti masih dapat dimodelkan dengan \hat{Y}_i . Langkah-langkah uji *Ramsey* sebagai berikut:

1. Dari persamaan (2) diperoleh \hat{Y}_i .
2. Dari persamaan (2), jika \hat{Y}_i dan \hat{u}_i masih punya pola berdasarkan plot, dapat diduga dengan menambahkan \hat{Y}_i dalam bentuk lain, seperti \hat{Y}_i^2 dan \hat{Y}_i^3 yang dapat meningkatkan nilai R^2 , sehingga terbentuk model seperti pada persamaan (3).

$$Y_i = \beta_1 + \beta_2 X_i + \beta_3 \hat{Y}_i^2 + \beta_4 \hat{Y}_i^3 + u_i \quad (3)$$

3. Setelah itu menghitung R^2 dari persamaan (2) dan (3). Kemudian menghitung uji F dengan persamaan (4).

$$F_{hit} = \frac{(R^2_{(model(3))} - R^2_{(model(2))})/k}{(1 - R^2_{(model(3))})/(n - p_2)} \quad (4)$$

Dengan $R^2_{(model(2))}$ adalah koefisien determinasi dari persamaan (2), $R^2_{(model(3))}$ adalah koefisien determinasi dari persamaan (3), k adalah jumlah prediktor tambahan, n adalah jumlah pengamatan, dan p_2 adalah banyaknya parameter dalam persamaan (3).

4. Jika $F_{hit} > F(\alpha, k, n - p_2)$ maka tolak H_0 . Pengujian hipotesis sebagai berikut:

H_0 : model persamaan (2) sesuai (model linier)

H_1 : model persamaan (2) tidak sesuai (model non linier)

Support Vector Machine merupakan metode untuk membuat prediksi baik pada kasus klasifikasi ataupun regresi (Santosa, 2007). Nugroho, *et al.* (2003) mengemukakan dengan sederhana konsep SVM, yaitu menemukan *hyperplane* terbaik yang dapat memisahkan dua kelas pada *input space*. *Hyperplane* berupa garis batas sebagai pemisah antara kedua kelas. *Hyperplane* yang letaknya di tengah-tengah antara dua set objek dari dua kelas disebut *hyperplane* terbaik. Jarak antara *hyperplane* dengan data terdekat yang berasal dari masing-masing kelas disebut *margin*. Data terdekat tersebut disebut sebagai *support vector*. Garis pemisah/*hyperplane* pada SVM yang dapat memisahkan data secara linier dapat dilihat pada persamaan (5) (Abe, 2010).

$$D(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + b \quad (5)$$

Penerapan SVM dalam kasus regresi disebut *Support Vector Regression* (Santosa, 2007). Menemukan garis pemisah atau *hyperplane* terbaik merupakan tujuan dari SVR (Maulana, *et al.*, 2019). Diberikan pasangan data *training* berukuran l , yaitu (\mathbf{x}_i, y_i) , $i = 1, \dots, l$ dengan data input berupa vektor, yaitu $\mathbf{x} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_l\} \subseteq \mathfrak{R}^m$ dan output berupa skalar, yaitu $y = \{y_1, \dots, y_l\} \subseteq \mathfrak{R}$. SVR akan menemukan suatu fungsi $f(x)$ yang mempunyai deviasi paling besar ε terhadap target aktual y_i . Ketika $\varepsilon = 0$ maka fungsi regresi akan sempurna. Fungsi regresi SVR diformulasikan pada persamaan (6) (Abe, 2010).

$$f(\mathbf{x}) = \mathbf{w}^T \varphi(\mathbf{x}) + b \quad (6)$$

Dengan \mathbf{w} merupakan vektor bobot, $\varphi(\mathbf{x})$ merupakan fungsi yang memetakan \mathbf{x} pada ruang berdimensi l , $l \geq m$, dan b adalah nilai konstanta bias.

Optimasi pemecahan masalah dalam bentuk *Quadratic Programming* seperti pada persamaan (7) (Abe, 2010).

$$\min \frac{1}{2} \|\mathbf{w}\|^2 \quad (7)$$

dengan batasan

$$\begin{aligned} y_i - \mathbf{w}^T \varphi(\mathbf{x}_i) - b &\leq \varepsilon, \text{ untuk } i = 1, \dots, l \\ \mathbf{w}^T \varphi(\mathbf{x}_i) + b - y_i &\leq \varepsilon, \text{ untuk } i = 1, \dots, l \end{aligned} \quad (8)$$

Persamaan (7) mengasumsikan seluruh titik berada dalam rentang $f \pm \varepsilon$ (*feasible*), jika beberapa titik berada di luar rentang $f \pm \varepsilon$ atau dalam kondisi ketidaklayakan (*infeasibility*) maka variabel *slack* ξ, ξ^* ditambahkan untuk menangani masalah pembatas yang tidak layak (*infeasible constraints*) dalam masalah optimasi. Masalah optimasi di atas kemudian diformulasikan sebagai berikut (Vapnik, dikutip dalam Santosa, 2007):

$$\min \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^l (\xi_i, \xi_i^*) \quad (9)$$

dengan batasan

$$\begin{aligned} y_i - \mathbf{w}^T \varphi(\mathbf{x}_i) - b - \xi_i &\leq \varepsilon \text{ untuk } i = 1, \dots, l \\ \mathbf{w}^T \varphi(\mathbf{x}_i) + b - y_i - \xi_i^* &\leq \varepsilon \text{ untuk } i = 1, \dots, l \\ \xi_i, \xi_i^* &\geq 0 \text{ untuk } i = 1, \dots, l \end{aligned} \quad (10)$$

Santosa (2007) mengemukakan bahwa konstanta $C > 0$ menentukan tawar-menawar (*trade off*) antara ketipisan fungsi (*flatness of function*) f dan batas atas deviasi lebih dari ε yang masih dapat ditoleransi. Semua deviasi yang lebih besar dari ε akan diberi penalti sebesar C . Penyelesaian optimasi untuk persamaan (9) dengan batasan pertidaksamaan (10) lebih mudah diselesaikan dengan mengubahnya menjadi formula *Lagrange* (Abe, 2010).

$$\begin{aligned} Q(\mathbf{w}, b, \xi, \xi^*, \alpha, \alpha^*, \eta, \eta^*) &= \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*) \\ &\quad - \sum_{i=1}^l \alpha_i (\varepsilon + \xi_i - y_i + \mathbf{w}^T \varphi(\mathbf{x}_i) + b) \\ &\quad - \sum_{i=1}^l \alpha_i^* (\varepsilon + \xi_i^* + y_i - \mathbf{w}^T \varphi(\mathbf{x}_i) - b) \end{aligned}$$

$$- \sum_{i=1}^l (\eta_i \xi_i + \eta_i^* \xi_i^*) \quad (11)$$

$\eta_i, \eta_i^*, \alpha_i, \alpha_i^*$ adalah *Lagrange Multiplier*. Solusi optimal dapat diperoleh dengan melakukan turunan parsial dari Q terhadap $\mathbf{w}, b, \xi, \xi^*$ sehingga \mathbf{w} ditulis sebagai (Abe, 2010):

$$\mathbf{w} = \sum_{i=1}^l (\alpha_i - \alpha_i^*) \varphi(\mathbf{x}_i) \quad (12)$$

optimal *hyperplane* ditulis sebagai berikut (Abe, 2010):

$$f(\mathbf{x}) = \sum_{i=1}^l (\alpha_i - \alpha_i^*) \varphi^T(\mathbf{x}_i) \varphi(\mathbf{x}) + b \quad (13)$$

misalkan $\beta_i = \alpha_i - \alpha_i^*$

maka $f(\mathbf{x}) = \sum_{i=1}^l \beta_i \varphi^T(\mathbf{x}_i) \varphi(\mathbf{x}) + b$

Solusi optimal untuk b dengan menggunakan KKT (*Karush Kuhn Tucker*) adalah (Abe, 2010):

$$\alpha_i (\varepsilon + \xi_i - y_i + \mathbf{w}^T \varphi(\mathbf{x}_i) + b) = 0 \text{ untuk } i = 1, \dots, l \quad (14)$$

$$\alpha_i^* (\varepsilon + \xi_i^* + y_i - \mathbf{w}^T \varphi(\mathbf{x}_i) - b) = 0 \text{ untuk } i = 1, \dots, l \quad (15)$$

$$\eta_i \xi_i = (C - \alpha_i) \xi_i = 0 \text{ untuk } i = 1, \dots, l \quad (16)$$

$$\eta_i^* \xi_i^* = (C - \alpha_i^*) \xi_i^* = 0 \text{ untuk } i = 1, \dots, l \quad (17)$$

Berdasarkan persamaan (16) karena $0 < \alpha_i < C$, maka $\xi_i = 0$, kemudian pada persamaan (17) karena $0 < \alpha_i^* < C$, maka $\xi_i^* = 0$. Pada persamaan (14) dan (15) akan menjadi persamaan (18) dan (19) (Abe, 2010):

$$\varepsilon - y_i + \mathbf{w}^T \varphi(\mathbf{x}_i) + b = 0 \text{ untuk } 0 < \alpha_i < C \quad (18)$$

$$\varepsilon + y_i - \mathbf{w}^T \varphi(\mathbf{x}_i) - b = 0 \text{ untuk } 0 < \alpha_i^* < C \quad (19)$$

Ini berarti bahwa data dengan residual $y - f(\mathbf{x}) = +\varepsilon$, α_i memenuhi $0 < \alpha_i < C$ dan data dengan residual $y - f(\mathbf{x}) = -\varepsilon$, α_i^* memenuhi $0 < \alpha_i^* < C$. Maka (Abe, 2010):

$$b = y_i - \mathbf{w}^T \varphi(\mathbf{x}_i) - \varepsilon \text{ untuk } 0 < \alpha_i < C \quad (20)$$

$$b = y_i - \mathbf{w}^T \varphi(\mathbf{x}_i) + \varepsilon \text{ untuk } 0 < \alpha_i^* < C \quad (21)$$

Data yang sulit dipisahkan secara linier dapat diatasi dengan metode kernel. Melalui fungsi pemetaan φ pada metode kernel, data \mathbf{x} pada *input space* dipetakan ke *feature space* F yang berdimensi lebih tinggi sehingga $\varphi : \mathbf{x} \rightarrow \varphi(\mathbf{x})$. Sering kali fungsi $\varphi(\mathbf{x})$ tidak tersedia dan tidak dapat dihitung namun dapat diselesaikan dengan *kernel trick*, yaitu menghitung *dot product* $\varphi(\mathbf{x}_i) \cdot \varphi(\mathbf{x}_j)$ di *feature space* yang digantikan fungsi kernel seperti pada persamaan (22) (Santosa, 2007).

$$K(\mathbf{x}_i, \mathbf{x}_j) = \varphi(\mathbf{x}_i) \cdot \varphi(\mathbf{x}_j) \quad (22)$$

Fungsi regresi SVR menjadi seperti persamaan (23).

$$f(\mathbf{x}) = \sum_{i,j=1}^l (\alpha_i - \alpha_i^*) K(\mathbf{x}_i, \mathbf{x}_j) + b \quad (23)$$

Fungsi kernel yang digunakan dalam penelitian ini adalah:

1. Kernel Linier

$$K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^T \cdot \mathbf{x}_j \quad (24)$$

2. Kernel *Polynomial*

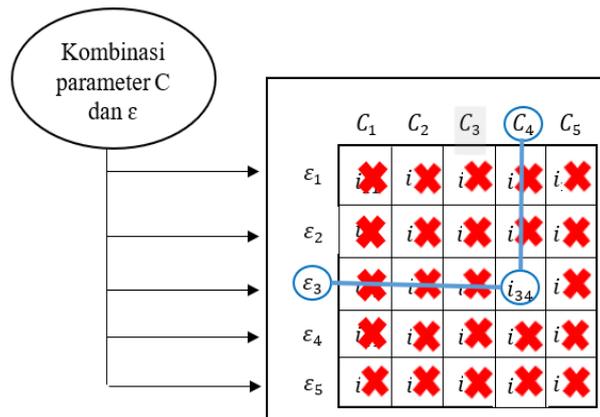
$$K(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i^T \cdot \mathbf{x}_j + 1)^d \quad (25)$$

3. Kernel *Radial Basis Function*

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2), \gamma > 0 \quad (26)$$

\mathbf{x}_i dan \mathbf{x}_j adalah pasangan dua data *training*. Parameter $d, \gamma > 0$ berupa konstanta (Fadilah, et al., 2020).

Hyperparameter yang optimal dapat ditentukan menggunakan metode *grid search* (Patriya, 2020).



Gambar 1. Ilustrasi Metode *Grid Search*

Metode *grid search* prosesnya hampir sama dengan pemeriksaan manual pada teknik *trial and error* yang memeriksa setiap kombinasi *hyperparameter* lalu membandingkan nilai *error* terkecil yang dihasilkan *hyperparameter* tersebut (Saputra, *et al.*, 2019). Algoritma *grid search* pada SVR kernel linier dapat dilakukan dengan cara:

1. Menentukan nilai *hyperparameter cost* (C) dan *epsilon* (ϵ). Nilai C yang digunakan yaitu 1 – 20; *epsilon* (ϵ) yaitu 0; 0,01; 0,02; 0,03; 0,04; 0,05; 0,06; 0,07; 0,08; 0,09; 0,1; 0,11; 0,12; 0,13; 0,14; 0,15; 0,16; 0,17; 0,18; 0,19; 0,2; 0,21; 0,22; 0,23; 0,24; 0,25; 0,26; 0,27; 0,28; 0,29; dan 0,3 (Saputra, *et al.*, 2019).
2. Membuat pasangan kombinasi nilai C dan ϵ , dengan semua nilai C harus berpasangan dengan ϵ begitupun sebaliknya.
3. Membuat model dengan metode SVR menggunakan masing-masing pasangan kombinasi nilai C dan ϵ .
4. Hitung *error* masing-masing kombinasi *hyperparameter*, kombinasi *hyperparameter* terbaik adalah pasangan *hyperparameter* dengan nilai *error* terendah.

Grid search menggunakan *cross validation* untuk melatih beberapa model (Patriya, 2020). *GridSearchCV* adalah metode untuk memilih kombinasi model dan *hyperparameter* yang menguji setiap kombinasi dan melakukan validasi untuk setiap kombinasi. Metode ini secara otomatis memvalidasi setiap kombinasi model dan *hyperparameter* sehingga dapat menghemat waktu pemrosesan (Hidayat, 2021). Pasangan *hyperparameter* yang menghasilkan akurasi terbaik dengan nilai *error* terkecil merupakan *hyperparameter* yang optimal (Purnama & Hendarsin, 2020). Penelitian ini menggunakan 10-fold CV dimana data latih dibagi menjadi 10 subset tanpa pengulangan. Proses iterasi dilakukan sebanyak 10 kali untuk mendapatkan 10 pengukuran akurasi. Setelah mendapatkan 10 pengukuran akurasi, dihitung rata-ratanya untuk mendapatkan kesalahan CV akhir dengan formula pada persamaan (27) (Wang, *et al.*, 2019).

$$CV_e = \frac{1}{10} \sum_{q=1}^{10} e_q \quad (27)$$

$$e_q = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2 \quad (28)$$

e_q adalah kesalahan rata-rata dari set pengujian ke- q dan n adalah jumlah data pengujian pada set pengujian ke- q .

Denormalisasi melakukan pengembalian data ke nilai awal setelah proses normalisasi untuk mendapatkan data asli. Rumus denormalisasi dalam rentang nilai [0, 1] seperti pada persamaan (29) (Maulana, *et al.*, 2019).

$$\text{denormalisasi}_i = x'_i(x_{max} - x_{min}) + x_{min} \quad (29)$$

Dengan *denormalisasi*_{*i*} adalah data hasil pengembalian ke nilai awal pada baris data ke-*i*, x'_i adalah data yang dinormalisasi pada baris data ke-*i*, x_{min} adalah nilai minimum data aktual, dan x_{max} adalah nilai maksimum data aktual.

Data *training* digunakan untuk melatih model dengan *hyperparameter* optimal agar menghasilkan model dengan kemampuan prediksi yang baik. Sedangkan data *testing* digunakan untuk pengujian model, sehingga model dapat menunjukkan prediksi yang diharapkan. Kemudian akan dihitung besarnya kesalahan hasil prediksi terhadap data aktual (Patriya, 2020).

Ukuran *error* yang digunakan dalam penelitian ini adalah:

1. *Mean Square Error* (MSE)

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (30)$$

2. *Symmetric Mean Absolute Percentage Error* (SMAPE)

$$SMAPE = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{(y_i + \hat{y}_i)/2} \cdot 100 \quad (31)$$

Sedangkan untuk menilai kecocokkan atau kesesuaian dari model prediksi, pada penelitian ini menggunakan koefisien determinasi (R^2). Koefisien determinasi merupakan ukuran untuk mengukur seberapa baik garis regresi sampel mencocokkan data. R^2 menunjukkan kesesuaian model, semakin besar R^2 semakin sesuai modelnya. R^2 dapat dijelaskan dengan perhitungan sebagai berikut (Gujarati, 1999):

$$R^2 = \frac{JKR}{JKT} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (32)$$

Dengan \hat{y}_i adalah nilai prediksi pada data ke-*i*, y_i adalah nilai output aktual pada data ke-*i*, dan \bar{y} adalah rata-rata nilai aktual.

3. METODE PENELITIAN

Jenis data yang digunakan dalam penelitian ini merupakan data sekunder, yaitu data historis saham harian perusahaan PT Anabatic Technologies Tbk (ATIC) berupa data *time series* mulai tanggal 01 Januari 2017 sampai tanggal 31 Desember 2021 sebanyak 1262 data dari situs resmi (www.finance.yahoo.com). Variabel yang digunakan dalam penelitian ini adalah harga penutupan saham harian perusahaan PT Anabatic Technologies Tbk (ATIC) yang dimodifikasi menjadi data ke-1,...,*N*-1 sebagai variabel prediktor (X) dan data ke-2,...,*N* sebagai variabel respon (Y). Pengolahan data dilakukan dengan *RStudio*. Langkah-langkah yang dilakukan pada penelitian ini adalah sebagai berikut:

1. Memasukkan data saham harian ATIC pada periode 01 Januari 2017 sampai tanggal 31 Desember 2021 pada *RStudio*.
2. Melakukan *preprocessing* data meliputi menghapus *missing value*, menghapus *outlier*, memodifikasi variabel prediktor (X) serta variabel respon (Y), dan normalisasi data.
3. Melakukan analisis deskriptif.
4. Membagi data menjadi data *training* dan data *testing* dengan proporsi tertentu.
5. Melakukan analisis *Support Vector Regression* (SVR).
6. Melakukan pengukuran kinerja dengan MSE, SMAPE, dan R^2 .
7. Melakukan *tuning hyperparameter* dengan *GridSearchCV* pada *Support Vector Regression* (SVR) untuk mendapatkan *hyperparameter* terbaik yang akan digunakan. Analisis data yang dilakukan adalah sebagai berikut:
 - a. Menentukan kernel dan nilai *hyperparameter* yang akan digunakan.
 - b. Melakukan analisis *Support Vector Regression* (SVR) setelah *tuning hyperparameter* dengan *hyperparameter* terbaik.
 - c. Melakukan pengukuran kinerja dengan MSE, SMAPE, dan R^2 .

8. Melakukan perbandingan hasil regresi *Support Vector Regression* (SVR) tanpa metode *GridSearchCV* dan dengan metode *GridSearchCV*.
9. Melakukan denormalisasi data.
10. Melakukan peramalan ke depan menggunakan model terbaik yang dihasilkan pada analisis SVR.

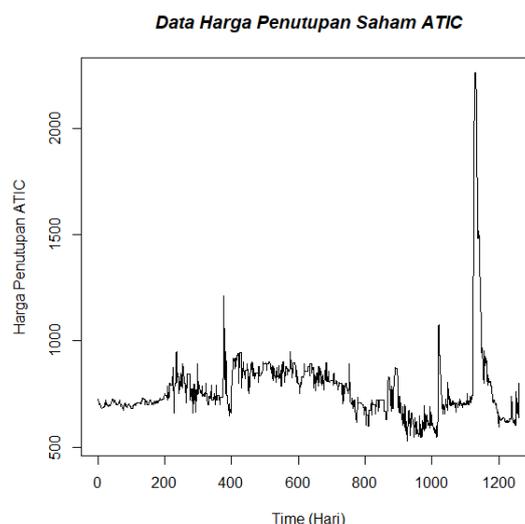
4. HASIL DAN PEMBAHASAN

Tahap *preprocessing* data yang dilakukan pada penelitian ini adalah pengecekan dan penanganan *missing value*, deteksi dan penanganan *outlier*, menentukan variabel X dan Y, dan normalisasi data. Setelah diamati terdapat 1 *missing value* dan 1 *outlier* ekstrem atau data yang berada paling jauh dari pengamatan lain sehingga dilakukan penghapusan pada data *missing value* dan *outlier* tersebut. Data menjadi sebanyak 1260 data. Penelitian ini menggunakan variabel *input* dan *output* karena metode yang akan digunakan termasuk ke dalam *supervised learning*. Data *input* yang digunakan merupakan variabel prediktor (X), yaitu harga penutupan saham ATIC satu periode sebelumnya yang digunakan untuk memprediksi harga penutupan saham satu hari selanjutnya. Pada masalah ini diasumsikan bahwa harga saham pada hari ini dipengaruhi oleh harga saham satu periode sebelumnya. Selanjutnya dilakukan normalisasi data dengan normalisasi *min-max* dimana data akan diubah ke dalam rentang yang lebih kecil, yaitu [0, 1]. Berikut adalah statistika deskriptif dari data harga penutupan saham harian ATIC mulai tanggal 01 Januari 2017 sampai tanggal 31 Desember 2021.

Tabel 1. Statistika Deskriptif

Data Harga Penutupan Saham Harian ATIC				
Min	Rata-rata	Maks	Varian	Std Dev
530	769,6	2260	27626,69	166,2128

Dari Tabel 1 diperoleh bahwa pada kurun waktu 01 Januari 2017 hingga 31 Desember 2021, rata-rata harga penutupan saham harian ATIC sebesar Rp 769,6/lembar dan standar deviasinya sebesar Rp 166,2128/lembar. Harga terendah pada harga penutupan ATIC adalah sebesar Rp 530/lembar yang terjadi pada tanggal 24 Agustus 2020 sedangkan harga tertinggi adalah sebesar Rp 2260/lembar yang terjadi pada tanggal 22 Juni 2021 hingga 28 Juni 2021. Berikut adalah grafik pergerakan harga penutupan saham ATIC mulai tanggal 01 Januari 2017 sampai tanggal 31 Desember 2021.



Gambar 2. Plot Data Harga Penutupan Saham ATIC

Pada Gambar 2 terlihat bahwa data harga penutupan saham ATIC menggambarkan kenaikan dan penurunan. Kenaikan dimulai sejak 15 Juni 2021 hingga 18 Juni 2021 yang mencapai 143%. Kenaikan yang signifikan tersebut menyebabkan perdagangan saham ATIC dihentikan sementara pada 21 Juni 2021. Saat diperdagangkan kembali pada 22 Juni 2021, harga penutupan saham ATIC meningkat hingga Rp 2260/lembar atau meningkat sejauh 24,86%. Peningkatan ini terjadi hingga 28 Juni 2021. Pada grafik juga terlihat bahwa harga penutupan saham ATIC sangat fluktuatif, artinya pola data tidak mengalami tren atau musiman.

Selanjutnya dilakukan uji non linieritas menggunakan uji *Ramsey RESET*. Berikut adalah hasil uji non linieritas dengan uji *Ramsey RESET*.

Tabel 2. Hasil Uji Non Linieritas *Ramsey RESET*

Data	Statistik Uji (<i>RESET</i>)	<i>p-value</i>
Harga Penutupan Saham ATIC	4,8757	0,007775

Berdasarkan Tabel 2 diperoleh bahwa nilai statistik uji *Ramsey RESET* sebesar 4,8757 lebih besar dari $F_{(0,05;2;1255)} = 3,0029$ atau nilai *p-value* sebesar 0,007775 lebih kecil dari $\alpha = 5\%$ maka dapat disimpulkan bahwa H_0 ditolak. Hal ini menunjukkan bahwa model non linier atau model yang dibentuk oleh variabel X dan Y membentuk hubungan non linier.

Kemudian dilakukan pembagian data *training* dan data *testing* dengan proporsi 80:20. Pembuatan model SVR dilakukan dengan dua cara, yaitu membentuk model SVR tanpa *GridSearchCV* dan dengan *GridSearchCV*. Model SVR tanpa *GridSearchCV* adalah model *default* yang tersedia pada *software RStudio* tanpa perlu mengubah kernel dan *hyperparameter* yang ada. Nilai *default* SVR yang tersedia pada *software RStudio* dengan *package e1017* adalah nilai *cost* (C) = 1, *gamma* (γ) = 1, dan *epsilon* (ϵ) = 0,1 dengan kernel RBF. Sedangkan pada model SVR dengan *GridSearchCV*, nilai *hyperparameter* harus ditentukan terlebih dahulu pada masing-masing kernel. *Hyperparameter* yang digunakan adalah *hyperparameter cost* (C), *epsilon* (ϵ), derajat (d) dan *gamma* (γ). Penelitian ini melakukan percobaan pada data sebanyak 10 kali dengan *trial and error* untuk tiap kombinasi *hyperparameter*.

Nilai-nilai *hyperparameter* yang diuji cobakan pada setiap model dan *hyperparameter* terbaik yang terpilih dapat dilihat pada Tabel 3.

Tabel 3. Nilai-nilai *Hyperparameter* Model SVR

Kernel	Nilai-nilai <i>Hyperparameter</i>				Model
	<i>cost</i> (C)	<i>epsilon</i> (ϵ)	derajat (d)	<i>gamma</i> (γ)	
Linier	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15,	0; 0,01; 0,02; 0,03; 0,04; 0,05; 0,06; 0,07; 0,08; 0,09; 0,1; 0,11; 0,12; 0,13; 0,14; 0,15; 0,16; 0,17; 0,18; 0,19; 0,2; 0,21; 0,22; 0,23; 0,24; 0,25; 0,26; 0,27; 0,28; 0,29; dan 0,3	- 1 dan 2 -	- - 0,0000002; 0,0000002; 0,000002; 0,0002; 0,002; 0,02; 0,2; 0; dan 2	$C = 1$ dan $\epsilon = 0,11$ $C = 2$; $\epsilon =$ 0,19; dan $d = 1$ $C = 1$; $\epsilon =$ 0,11; dan $\gamma = 0,02$

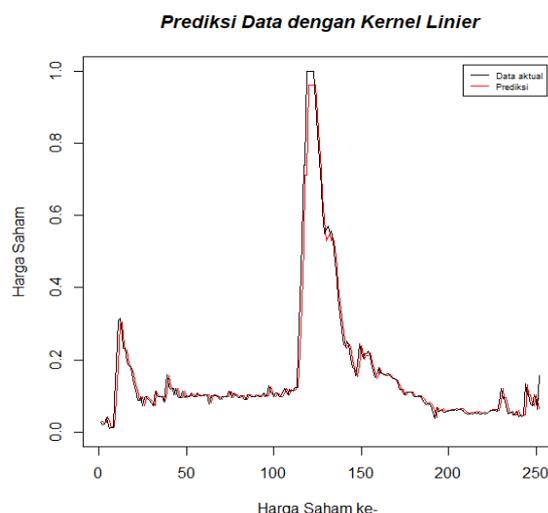
Nilai *hyperparameter* tersebut dipilih berdasarkan hasil percobaan yang telah dilakukan. Nilai *epsilon* (ϵ) yang kecil membuat batas toleransi *error* yang dibatasi kecil sehingga kesalahan prediksi akan semakin kecil. Sebaliknya, nilai *epsilon* (ϵ) yang besar membuat batas toleransi *error* yang dibatasi besar sehingga kesalahan prediksi akan semakin besar. Hasil evaluasi model SVR tanpa *GridSearchCV* dan dengan *GridSearchCV* dapat dilihat pada Tabel 4.

Tabel 4. Hasil Evaluasi Model

Model SVR	MSE	SMAPE	R^2
SVR <i>default</i>	0,02993634	0,2319903	0,1545886
SVR kernel linier	0,001237173	0,1167301	0,9206643
SVR kernel <i>polynomial</i>	0,001356706	0,12674	0,8480179
SVR kernel <i>radial basis function</i>	0,0290428	0,2056089	0,1587606

Berdasarkan Tabel 4, model terbaik dilihat dari R^2 yang besar serta MSE dan SMAPE yang kecil. Diperoleh bahwa model SVR kernel linier lebih baik dan lebih akurat dalam melakukan prediksi harga penutupan saham ATIC dibandingkan dengan model SVR *default*, SVR *polynomial* maupun SVR kernel RBF. Perhitungan kesalahan prediksi, akurasi, dan kesesuaian model SVR kernel linier menghasilkan MSE rendah sebesar 0,001237173; SMAPE sebesar 0,1167301 artinya model mempunyai kemampuan prediksi yang akurat; dan $R^2 = 0,9206643$ artinya model prediksi yang dibentuk dengan metode SVR kernel linier sesuai untuk melakukan prediksi data harga penutupan saham harian ATIC.

Plot hasil prediksi dan data aktual menggunakan metode SVR kernel linier dapat dilihat pada Gambar 3.



Gambar 3. Plot Data Prediksi Model SVR Kernel Linier

Gambar 3 menunjukkan bahwa prediksi dengan SVR kernel linier mengikuti pola data yang ada sehingga prediksi yang dilakukan menjadi lebih sesuai dengan data aktual atau memperlihatkan bahwa data aktual dibandingkan dengan data hasil prediksi memiliki pola yang hampir sama, sehingga dapat dikatakan bahwa model SVR dengan kernel linier dapat mengatasi masalah *overfitting* dalam memprediksi harga saham ATIC.

Model terbaik menggunakan persamaan SVR yang telah terbentuk dapat digunakan untuk meramalkan harga penutupan saham ATIC pada waktu yang akan datang. Pada tahap ini dilakukan peramalan harga penutupan saham ATIC untuk satu bulan ke depan atau 30 hari di bulan Januari 2022. Berikut hasil prediksi untuk tanggal 01 Januari 2022 sampai dengan 30 Januari 2022 yang didapat dari model SVR kernel linier dengan *hyperparameter*

terbaik setelah proses *GridSearchCV*, yaitu $C = 1$ dan $\varepsilon = 0,11$ serta pembagian data *training* dan *testing* sebesar 80:20.

Tabel 5. Peramalan Harga Penutupan Saham ATIC pada 01 Januari 2022 – 30 Januari 2022

Tanggal	Peramalan	Tanggal	Peramalan
01/01/2022	623,9582	16/01/2022	548,4523
02/01/2022	608,3063	17/01/2022	547,9016
03/01/2022	595,7848	18/01/2022	547,4610
04/01/2022	585,7676	19/01/2022	547,1086
05/01/2022	577,7538	20/01/2022	546,8266
06/01/2022	571,3428	21/01/2022	546,6010
07/01/2022	566,2140	22/01/2022	546,4206
08/01/2022	562,1110	23/01/2022	546,2762
09/01/2022	558,8285	24/01/2022	546,1607
10/01/2022	556,2026	25/01/2022	546,0683
11/01/2022	554,1018	26/01/2022	545,9944
12/01/2022	552,4212	27/01/2022	545,9353
13/01/2022	551,0767	28/01/2022	545,8880
14/01/2022	550,0011	29/01/2022	545,8501
15/01/2022	549,1406	30/01/2022	545,8199

Tabel 5 menunjukkan bahwa peramalan harga penutupan saham ATIC dalam 30 hari ke depan. Berdasarkan hasil tersebut diperoleh bahwa harga penutupan saham ATIC mengalami penurunan dari hari ke hari.

5. KESIMPULAN

Penerapan metode SVR tanpa *GridSearchCV* menghasilkan nilai MSE yang agak tinggi, yaitu 0,02993634 dan SMAPE yang menunjukkan prediksi kurang akurat, yaitu 0,2319903 atau 23,19903%; dan nilai R^2 sebesar 0,1545886 atau 15,45886% artinya model prediksi yang dibentuk dengan metode SVR tanpa *GridSearchCV* tidak sesuai untuk melakukan prediksi data harga penutupan saham harian ATIC. Penentuan *hyperparameter* yang optimal menggunakan *grid search cross validation* dapat dilakukan agar diperoleh model prediksi yang sesuai untuk melakukan prediksi data harga penutupan saham harian ATIC. Model prediksi dibuat berdasarkan *hyperparameter* optimal yang diperoleh dari *tuning hyperparameter* menggunakan *GridSearchCV*. Model terbaik yang diperoleh, yaitu model SVR kernel linier dengan $C = 1$ dan $\varepsilon = 0,11$. Berdasarkan hasil perhitungan evaluasi kinerja, model SVR dengan kernel linier memiliki R^2 yang paling besar serta memiliki nilai MSE dan SMAPE paling kecil dibandingkan dengan model SVR dengan kernel *polynomial*, model SVR dengan kernel *radial basis function*, dan model SVR tanpa *GridSearchCV*. Dengan model terbaik, selanjutnya dilakukan peramalan harga penutupan saham ATIC dalam 30 hari ke depan dan diperoleh hasil bahwa harga penutupan saham ATIC cenderung mengalami penurunan dari hari ke hari.

DAFTAR PUSTAKA

- Abe, S. 2010. *Support Vector Machines for Pattern Classification*. London: Springer-Verlag.
- Antony, E., Sreekanth, N. S., Kumar, R. K. S. & T, N. 2021. *Data Preprocessing Techniques for Handling Time Series data for Environmental Science Studies*. *International Journal of Engineering Trends and Technology*, 69(5), pp. 196-207.

- Bursa Efek Indonesia. 2018. *Saham*. <https://www.idx.co.id/produk/saham/>
Diakses: 21 Februari 2022.
- Fadilah, W. R. U., Agfiannisa, D. & Azhar, Y. 2020. *Analisis Prediksi Harga Saham PT. Telekomunikasi Indonesia Menggunakan Metode Support Vector Machine*. *Fountain of Informatics Journal*, 5(2), pp. 45-51.
- Gujarati, D. 1999. *Ekonometrika Dasar*. Jakarta: Erlangga.
- Han, J., Kamber, M. & Pei, J. 2012. *Data Mining Concepts and Techniques*. USA: Elsevier.
- Hidayat, M. A. 2021. *GridSearchCV*.
<https://algotech.netlify.app/blog/gridsearchcv/#:~:text=GridSearchCV%20merupakan%20bagian%20dari%20modul,masing%20secara%20otomatis%20dan%20sistematis>. Diakses: 22 Februari 2022.
- Intan, K. 2021. *Kembali Diperdagangkan, Saham Anabatic Technologies (ATIC) Melesat 24,86%*. <https://investasi.kontan.co.id/news/kembali-diperdagangkan-saham-anabatic-technologies-atic-melesat-2486>. Diakses: 21 Februari 2022.
- Makridakis, S., Wheelwright, S. C. & McGee, V. E. 1992. *Metode dan Aplikasi Peramalan*. Jakarta: Erlangga.
- Maulana, N. D., Setiawan, B. D. & Dewi, C. 2019. *Implementasi Metode Support Vector Regression (SVR) Dalam Peramalan Penjualan Roti (Studi Kasus: Harum Bakery)*. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 3(3), pp. 2986-2995.
- Nugroho, A. S., Witarto, A. B. & Handoko, D. 2003. *Support Vector Machine: Teori dan Aplikasinya dalam Bioinformatika*. Japan, Proceeding of Indonesian Scientific Meeting.
- Patriya, E. 2020. *Implementasi Support Vector Machine Pada Prediksi Harga Saham Gabungan (IHSG)*. *Jurnal Ilmiah Teknologi dan Rekayasa*, 25(1), pp. 24-38.
- Prabowo, H., Suhartono & Prastyo, D. D. 2020. *The Performance of Ramsey Test, White Test and Terasvirta Test in Detecting Nonlinearity*. *INFERENSI*, 3(1), pp. 1-12.
- Purnama, D. I. & Hendarsin, O. P. 2020. *Peramalan Jumlah Penumpang Berangkat Melalui Transportasi Udara di Sulawesi Tengah Menggunakan Support Vector Regression (SVR)*. *Jambura Journal of Mathematics*, Juli, 2(2), pp. 49-59.
- Santosa, B. 2007. *Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Yogyakarta: Graha Ilmu.
- Saputra, G. H., Wigena, A. H. & Sartono, B. 2019. *Penggunaan Support Vector Regression dalam Pemodelan Indeks Saham Syariah Indonesia dengan Algoritme Grid Search*. *Indonesian Journal of Statistics and Its Applications*, 3(2), pp. 148-160.
- Wang, X., Gong, G. & Li, N. 2019. *Automated Recognition of Epileptic EEG States Using Automated Recognition of Epileptic EEG States Using, Gradient Boosting Machine, and Grid Search Optimizer*. *Sensors*, 19(219), pp. 1-18.