

IMPLEMENTASI METODE *SUPPORT VECTOR REGRESSION* UNTUK PREDIKSI HARGA SAHAM PT. ADARO ENERGY TBK

Anik Sri Mulyani^{1*}, Mustafid², Tatik Widiharah³

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

*e-mail: aniksrimulyani@students.undip.ac.id

DOI: 10.14710/j.gauss.14.2.457-468

Article Info:

Received: 2024-07-28

Accepted: 2025-11-06

Available Online: 2025-11-19

Keywords:

Stock; Support Vector Regression;
Grid search; Kernels.

Abstract: Stock investment is very attractive to investors because it provides many benefits. The higher the profit offered in investing, the higher the risk investors will face. Stock price analysis is needed to predict stock prices to reduce the risk of loss. Stock data has dynamic, non-linear, and unpredictable characteristics, so a method is required to overcome the limitations of time series data. The Support Vector Regression method will be applied in this study to predict PT. Adaro stock prices. The Support Vector Regression method is one method that does not require assumptions, so it can be used to overcome the limitations of regression analysis with time series data. The problem often faced when using the SVR method is determining the optimal hyperparameters. This study determines the optimal hyperparameter using the grid search algorithm. The data is divided into training and testing data with a ratio of 90:10. The best kernel to predict the share price of PT. Adaro is using a linear kernel with a value of Cost = 4 and epsilon = 0.0001. The model produces MAPE testing data of 1.608%, which means the model is very good for prediction.

1. PENDAHULUAN

Kinerja pasar modal berkembang positif selama tahun 2022, sehingga atensi masyarakat untuk berinvestasi di pasar modal Indonesia turut meningkat. Investor memiliki permintaan yang tinggi terhadap investasi saham, baik investasi saham jangka pendek ataupun jangka panjang (Yudhawan dan Purwaningsih, 2020). Investasi saham menawarkan banyak profit bagi investor. Semakin tinggi profit yang ditawarkan dalam berinvestasi, semakin tinggi pula resiko yang akan dihadapi. Data saham memiliki karakteristik yang dinamis, non linier, dan tidak mudah diprediksi (Vijh *et al.*, 2020). Investor cenderung membeli saham yang memiliki kinerja dan prospek yang baik untuk menekan resiko kerugian. PT. Adaro Energy Tbk memiliki saham ADRO dengan rasio keuangan yang sehat dan prospek yang dianggap positif, terutama seiring dengan kenaikan harga batu bara di pasar global. Harga saham sangat fluktuatif, sehingga sulit bagi investor untuk memprediksi harga saham. Teknologi yang semakin maju seperti sekarang dapat membantu investor memperkirakan harga saham dan membantu dalam pengambilan keputusan investasi (Fitrinanda dan Djunaedi, 2022). Harga saham dicatat selama periode tertentu, sehingga dapat dianalisis dan dimodelkan dengan menggunakan analisis runtun waktu. Metode ARIMA mempunyai beberapa asumsi yang harus dipenuhi, diantaranya yaitu mengikuti pola stasioner (konstan), residual berdistribusi normal, dan memenuhi kondisi *white noise*. Apabila asumsi-asumsi tersebut tidak terpenuhi maka metode ARIMA tidak dapat digunakan sehingga kehadiran *Support Vector Regression* (SVR) menjadi salah satu metode alternatif untuk mengatasi hal tersebut.

Support Vector Regression (SVR) merupakan salah satu *machine learning* yang bisa diaplikasikan untuk memprediksi data *time series* (Lai & Liu, 2010). Ide dibalik SVR adalah mengoptimalkan *hyperplane* untuk memperoleh data-data yang akan menjadi *support vector*. Keuntungan dari SVR adalah dapat mengatasi *overfitting*. Selain itu, metode SVR

juga merupakan salah satu metode bebas asumsi yang dapat mengatasi keterbatasan pada analisis regresi dengan data *time series*. SVR memiliki beberapa kernel, diantaranya seperti: linier, polynomial, RBF, dan sigmoid. Masalah yang sering dihadapi saat menggunakan metode *Support Vector Regression* (SVR) adalah penentuan *hyperparameter* untuk model yang optimal. Metode *grid search* merupakan salah satu metode optimasi yang dapat digunakan untuk menentukan model optimal (Yasin *et al.*, 2014). Penelitian ini akan menerapkan metode *Support Vector Regression* untuk prediksi saham ADRO agar dapat membantu investor dalam mengambil tindakan pada saat berinvestasi.

2. TINJAUAN PUSTAKA

Saham dapat diartikan sebagai tanda bukti kepemilikan atau penyertaan modal individu atau badan usaha dalam sebuah perusahaan. Pemilik saham berhak menerima *dividen* dan hak atas pendapatan perusahaan serta asset perusahaan. Saham dapat diperjualbelikan dengan mudah, sehingga termasuk kedalam harta lancar. Nilai moneter dari saham yang ditentukan dari aksi pembelian dan penawaran saham oleh anggota bursa di bursa efek disebut harga saham (Hadi, 2013).

Machine Learning adalah kumpulan metode komputasi yang mampu meningkatkan kinerja melalui pemanfaatan pengetahuan yang berasal dari pengalaman (Heryadi dan Wahyono, 2020). Terdapat dua pendekatan *machine learning* yaitu *unsupervised learning* dan *supervised learning*. *Unsupervised Learning* merupakan metode pembelajaran mempergunakan sejumlah contoh data tanpa label, contohnya *clustering*. Sedangkan *Supervised Learning* merupakan metode pembelajaran yang bertujuan untuk memprediksi variabel target sebuah data baru berdasarkan model hasil *training* menggunakan himpunan data berlabel, contohnya regresi (Santoso, 2013).

Normalisasi data adalah sebuah proses pengolahan data dengan tujuan untuk mengubah data agar berada dalam kisaran nilai tertentu sehingga setiap data berada dalam jangkauan yang sama. Normalisasi data juga bertujuan untuk mengurangi waktu komputasi program. Salah satu metode untuk normalisasi data yaitu *min-max normalization*. Rumus umum *min-max normalization*:

$$X' = \frac{(X - X_{min})}{(X_{max} - X_{min})} (X_{max\ baru} - X_{min\ baru}) + X_{min\ baru} \quad (1)$$

Persamaan (2) merupakan rumus *Min-max normalization* untuk mengubah nilai data sehingga berada pada rentang antara 0 sampai 1.

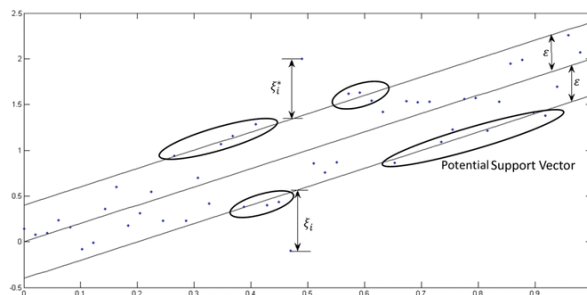
$$X' = \frac{(X - X_{min})}{(X_{max} - X_{min})} \quad (2)$$

Denormalisasi data merupakan suatu proses yang digunakan untuk mengembalikan nilai-nilai data yang sudah dinormalisasi kedalam nilai aslinya. Denormalisasi dihitung menggunakan persamaan (3).

$$X = X_{min} + X'(x_{max} - x_{min}) \quad (3)$$

dengan X' merupakan data setelah normalisasi, X merupakan data asli/aktual, X_{min} merupakan nilai data paling kecil, dan X_{max} merupakan nilai data paling besar (Priliani *et al.*, 2018).

Support Vector Regression (SVR) adalah perkembangan dari *Support Vector Machine* (SVM) yang dimanfaatkan khusus untuk kasus regresi. Teknik SVR mempunyai kemampuan dalam memaksimalkan akurasi prediksi dan menghindari *overfitting* pada saat yang sama (Mahdevari *et al.*, 2014). Gambar 1 menunjukkan sebuah *hyperplane* dan dua garis atas dan bawah sebagai pembatas, ε merupakan jarak antara *hyperplane* dan garis pembatas, dan data yang dilingkari merupakan data yang berpotensi menjadi *support vector*.



Gambar 1. Support Vector Regression
(Sumber: Muthiah, 2022)

Ide dasar pada metode SVR dapat disimpulkan data *training* $\{(\mathbf{x}_i, y_i)\}, i = 1, 2, \dots, l$ dan l adalah banyaknya data *training*. Vektor input $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{id}) \in R^d$ dan output $y_i \in R$. Persamaan (4) merupakan fungsi regresi sebagai *optimal hyperplane* dengan metode SVR (Santoso, 2013).

$$f(\mathbf{x}) = \mathbf{w}^T \boldsymbol{\varphi}(\mathbf{x}) + b \quad (4)$$

dengan $\mathbf{w}^T = (w_1, w_2, \dots, w_d)$ merupakan vektor bobot, $\mathbf{x}^T = (x_1, x_2, \dots, x_d)$ merupakan vektor input, $\boldsymbol{\varphi}(\mathbf{x})$ merupakan representasi suatu titik dalam *feature space* yang dihasilkan dari pemetaan x dalam *input space* dan b merupakan intercept.

Dalam rangka mencapai generalisasi yang optimal diperlukan cara agar $\mathbf{w}$ minimal. Persamaan (5) merupakan meminimalkan $\|\mathbf{w}\|^2$ yang diselesaikan dengan bentuk *Quadratik Programming*

$$\min \frac{1}{2} \|\mathbf{w}\|^2 \quad (5)$$

dengan syarat :

$$\begin{aligned} y_i - \mathbf{w}^T \boldsymbol{\varphi}(\mathbf{x}_i) - b &\leq \varepsilon \\ \mathbf{w}^T \boldsymbol{\varphi}(\mathbf{x}_i) + b - y_i &\leq \varepsilon \end{aligned} \quad (6)$$

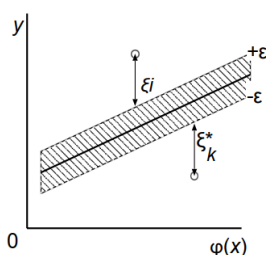
dengan y_i merupakan nilai aktual data ke i . faktor $\|\mathbf{w}\|^2$ disebut dengan regulasi. $\|\mathbf{w}\|^2$ juga dapat dituliskan dengan $\mathbf{w}^T \mathbf{w}$

Dalam persamaan (5), diasumsikan bahwa seluruh titik berada dalam rentang $f(\mathbf{x}) \pm \varepsilon$. Namun, jika nilai kesalahan melebihi ambang ε , yang disebabkan oleh titik-titik diluar rentang $f(\mathbf{x}) \pm \varepsilon$, maka diperlukan *soft margin* atau *variable slack* ($\xi_i + \xi_i^*$) sehingga persamaannya berubah menjadi:

$$\min \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \quad (7)$$

dengan syarat :

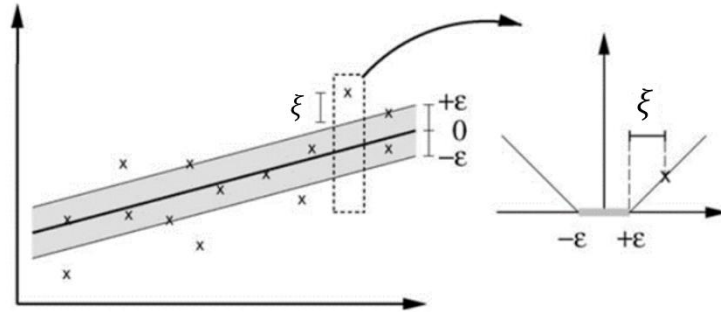
$$\begin{aligned} y_i - \mathbf{w}^T \boldsymbol{\varphi}(\mathbf{x}_i) - b &\leq \varepsilon + \xi_i \\ \mathbf{w}^T \boldsymbol{\varphi}(\mathbf{x}_i) + b - y_i &\leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* &\geq 0 \end{aligned} \quad (8)$$



Gambar 2. Variable Slack
(Sumber: Abe, 2010)

Pada gambar 2, ξ merupakan variable *slack* yang dapat dihitung dengan $|y - f(x)| - \varepsilon$. Nilai ε yang rendah berarti variabel *slack* memiliki nilai yang tinggi, sehingga dapat meningkatkan akurasi, dan sebaliknya. Nilai *slack* yang berlebihan dapat menyebabkan kesalahan empiris dalam perhitungan (Smola dan Scholkopf, 2004).

Loss function adalah sebuah fungsi yang menggambarkan hubungan antara kesalahan dengan besaran pinalti yang dikenakan. Penggunaan *Loss function* yang berbeda akan mendapatkan formula SVR yang berbeda pula (Santoso, 2013). Salah satu *Loss function* yang dapat digunakan adalah ε - insensitive loss function.



Gambar 3. SVR dengan ε -insensitive loss function
(Sumber: Smola dan Scholkopf, 2004)

Konsep dari ε -insensitive Loss function dijelaskan secara grafis pada Gambar 3. Hanya sampel di luar *margin* ε yang akan memiliki variabel slak bernilai positif. Biasanya, jika nilai prediksi berada di dalam wilayah ε , maka kerugian akan menjadi nol, sedangkan jika titik prediksi berada di luar *margin* ε , kerugian akan menjadi nilai perbedaan antara nilai prediksi dan ε (Mahdevari *et al.*, 2014). Persamaan (9) merupakan formulasi ε - insensitive loss function

$$L_{\varepsilon} = \begin{cases} 0 & \text{jika } |f(x) - y| \leq \varepsilon \\ |f(x) - y| - \varepsilon & \text{lainnya} \end{cases} \quad (9)$$

Persamaan (7) dengan batas persamaan (8) dapat diselesaikan menggunakan persamaan *Lagrange Multiplier* :

$$L = \left(\frac{1}{2} (\mathbf{w}^T \mathbf{w}) + C \sum_{i=1}^l (\xi_i + \xi_i^*)\right) - \left(\sum_{i=1}^l \alpha_i (\varepsilon + \xi_i - y_i + \mathbf{w}^T \boldsymbol{\varphi}(\mathbf{x}_i) + b)\right) - \left(\sum_{i=1}^l \alpha_i^* (\varepsilon + \xi_i^* + y_i - \mathbf{w}^T \boldsymbol{\varphi}(\mathbf{x}_i) - b)\right) - \left(\sum_{i=1}^l (\eta_i \xi_i + \eta_i^* \xi_i^*)\right) \quad (10)$$

dengan $C > 0$ dan $\alpha_i, \alpha_i^*, \eta_i, \eta_i^*$ merupakan *lagrange multiplier*. Kemudian dilakukan turunan parsial dari L terhadap $\mathbf{w}, b, \xi_i$, dan ξ_i^* agar mendapatkan solusi yang optimal.

$$\frac{\partial L}{\partial \mathbf{w}} = \frac{\partial L}{\partial w_1}, \frac{\partial L}{\partial w_2}, \dots, \frac{\partial L}{\partial w_d} \quad (11)$$

$$\frac{\partial L}{\partial w_1} = w_1 - \sum_{i=1}^l (\alpha_i - \alpha_i^*) \boldsymbol{\varphi}(\mathbf{x}_i) \quad (12)$$

$$\frac{\partial L}{\partial w_2} = w_2 - \sum_{i=1}^l (\alpha_i - \alpha_i^*) \boldsymbol{\varphi}(\mathbf{x}_i) \quad (13)$$

$$\frac{\partial L}{\partial w_d} = w_d - \sum_{i=1}^l (\alpha_i - \alpha_i^*) \boldsymbol{\varphi}(\mathbf{x}_i) \quad (14)$$

$$\frac{\partial L}{\partial \mathbf{w}} = \mathbf{w} - \sum_{i=1}^l (\alpha_i - \alpha_i^*) \boldsymbol{\varphi}(\mathbf{x}_i) \quad (15)$$

$$\frac{\partial L}{\partial b} = \sum_{i=1}^l (\alpha_i^* - \alpha_i) \quad (16)$$

$$\frac{\partial L}{\partial \xi_i} = C - \alpha_i - \eta_i \quad (17)$$

$$\frac{\partial L}{\partial \xi_i^*} = C - \alpha_i^* - \eta_i^* \quad (18)$$

dengan syarat ekstrem $\frac{\partial L}{\partial \mathbf{w}} = 0$, $\frac{\partial L}{\partial b} = 0$, $\frac{\partial L}{\partial \xi_i} = 0$, dan $\frac{\partial L}{\partial \xi_i^*} = 0$ diperoleh:

$$\mathbf{w} = \sum_{i=1}^l (\alpha_i - \alpha_i^*) \boldsymbol{\varphi}(\mathbf{x}_i) \quad (19)$$

$$\sum_{i=1}^l (\alpha_i^* - \alpha_i) = 0 \quad (20)$$

$$C = \alpha_i + \eta_i \quad (21)$$

$$C = \alpha_i^* + \eta_i^* \quad (22)$$

Solusi optimasi pada persamaan (10) dengan mensubstitusikan persamaan (19), (20), (21), dan (22) didapat persamaan (23)

$$\begin{aligned} L(\alpha_i, \alpha_i^*) = & -\frac{1}{2} \sum_{i=1, j=1}^l (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) \boldsymbol{\varphi}^T(\mathbf{x}_i) \boldsymbol{\varphi}(\mathbf{x}_j) - \varepsilon \sum_{i=1}^l (\alpha_i + \alpha_i^*) \\ & + \sum_{i=1}^l y_i (\alpha_i - \alpha_i^*) \end{aligned} \quad (23)$$

dengan syarat:

$$\begin{aligned} \sum_{i=1}^l (\alpha_i - \alpha_i^*) &= 0, \\ 0 \leq \alpha_i &\leq C \quad i = 1, 2, \dots, l \\ 0 \leq \alpha_i^* &\leq C \quad i = 1, 2, \dots, l \end{aligned} \quad (24)$$

Nilai b dimaksimalkan dari kondisi dengan menggunakan *Karush-Kuhn-Tucker* (KKT):

$$\alpha_i (\varepsilon + \xi_i - y_i + \mathbf{w}^T \boldsymbol{\varphi}(\mathbf{x}_i) + b) = 0 \quad (25)$$

$$\alpha_i^* (\varepsilon + \xi_i^* + y_i - \mathbf{w}^T \boldsymbol{\varphi}(\mathbf{x}_i) - b) = 0 \quad (26)$$

$$(C - \alpha_i) \xi_i = 0 \quad (27)$$

$$(C - \alpha_i^*) \xi_i^* = 0 \quad (28)$$

Pada persamaan (27) ketika $0 < \alpha_i < C$ maka $\xi_i = 0$, dan persamaan (28) ketika $0 < \alpha_i^* < C$ maka $\xi_i^* = 0$. Jadi persamaan (25) dan (26) diperoleh b :

$$b = y_i - \mathbf{w}^T \boldsymbol{\varphi}(\mathbf{x}_i) - \varepsilon \quad \text{untuk } 0 < \alpha_i < C \quad (29)$$

$$b = y_i - \mathbf{w}^T \boldsymbol{\varphi}(\mathbf{x}_i) + \varepsilon \quad \text{untuk } 0 < \alpha_i^* < C \quad (30)$$

Pada persamaan (4) dan (19) fungsi $f(\mathbf{x})$ dapat ditulis:

$$f(\mathbf{x}) = \sum_{i=1}^l (\alpha_i - \alpha_i^*) \boldsymbol{\varphi}^T(\mathbf{x}_i) \boldsymbol{\varphi}(\mathbf{x}) + b \quad (31)$$

Suatu fungsi kernel $K(\mathbf{x}_i, \mathbf{x})$ bisa digunakan untuk menggantikan *dot-product* antara dua vektor $(\boldsymbol{\varphi}^T(\mathbf{x}_i) \cdot \boldsymbol{\varphi}(\mathbf{x}))$.

Fungsi regresi $f(x)$ secara eksplisit dapat dirumuskan sebagai berikut:

$$f(x) = \sum_{i=1}^l (\alpha_i - \alpha_i^*) K(\mathbf{x}_i, \mathbf{x}) + b \quad (32)$$

Beberapa fungsi kernel yang sering digunakan dalam aplikasi:

- a. Kernel linier, dinotasikan sebagai berikut:

$$K(\mathbf{x}_i, \mathbf{x}) = \mathbf{x}_i^T \mathbf{x} \quad (33)$$

- b. Kernel polynomial

$$K(\mathbf{x}_i, \mathbf{x}) = (\mathbf{x}_i^T \mathbf{x} + c)^d \quad (34)$$

- c. Kernel *Radial Basis Function* (RBF)

$$K(\mathbf{x}_i, \mathbf{x}) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}\|^2) \quad (35)$$

- d. Kernel sigmoid

$$K(\mathbf{x}_i, \mathbf{x}) = \tanh(\gamma(\mathbf{x}_i^T \mathbf{x}) + c) \quad (36)$$

Nilai optimal dari suatu model atau fungsi dapat dicari menggunakan metode *grid search* dengan mencari kombinasi *hyperparameter* yang memberikan nilai terbaik (dalam hal ini, memiliki *error* yang paling kecil) (Raschka dan Mirjalali, 2019). *Grid search* mengoptimalkan SVR menggunakan teknik *cross validation* sebagai metrik kinerja. Dalam proses pengolahan data, *cross-validation* digunakan untuk membagi data *training* menjadi beberapa *fold* yang terdiri dari data pelatihan dan data validasi. Tiap model yang terbentuk dari berbagai kombinasi *hyperparameter* kemudian menjalani *cross-validation* untuk menghitung rata-rata skor MSE dari seluruh *fold* yang terlibat. Teknik *cross validation* untuk data *time series* dapat diterapkan dengan *library Time Series Split* dan *Grid search CV* yang tersedia dalam *scikit-learn*. Dengan cara ini, optimasi *grid search* dapat diterapkan dengan menggunakan teknik *time series 5-fold cross-validation* (Patriya, 2020).

Pada tahun 1939 Tinbergen memulai perkembangan teknik evaluasi akurasi metode peramalan, sebagai tanggapan terhadap Keynes, yang menyatakan bahwa teori-teori harus dikonfirmasi jika data dan metode statistik bekerja dengan baik (Woschnagg dan Cipan, 2004). Untuk mengkonfirmasinya dibutuhkan evaluasi model. Evaluasi model dapat dilakukan dengan cara menghitung MSE dan MAPE sebagai berikut:

1. MSE (*Mean Square Error*)

Dalam penelitian ini, MSE digunakan sebagai dasar pemilihan kombinasi *hyperparameter* terbaik pada proses *grid search*. MSE dirumuskan dengan:

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - f(x_i))^2 \quad (37)$$

2. MAPE (*Mean Absolute Percentage Error*)

MAPE dapat digunakan sebagai evaluasi proses prediksi dan ukuran akurasi yang ditunjukkan dalam presentase. Dalam penelitian ini, MAPE juga digunakan untuk pemilihan model dengan kernel terbaik. MAPE dapat dihitung dengan:

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \left(\frac{y_i - f(x_i)}{y_i} \right) \right| \times 100\% \quad (38)$$

3. METODE PENELITIAN

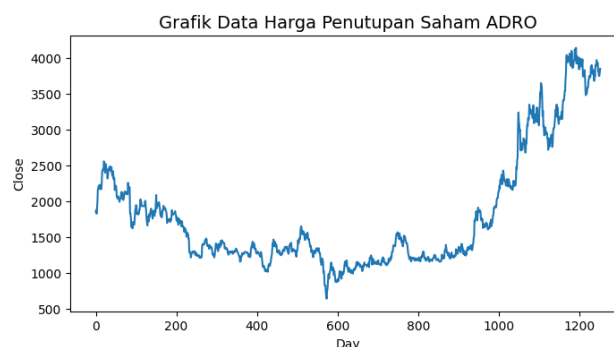
Pada penelitian ini menggunakan data penutupan harga saham harian (*closing price*) PT. Adaro Energy Tbk pada periode 1 Januari 2018 - 31 Desember 2022 sebanyak 1253 data. Data didapat dari Situs Yahoo finance (<http://finance.yahoo.com>).

Tahapan analisis dalam penelitian ini meliputi:

1. Mempersiapkan data saham harian PT. Adaro Energy Tbk
2. Melakukan statistika deskriptif terhadap data harga penutupan saham harian yang sudah dipersiapkan
3. Normalisasi data
4. Menentukan variabel *dependent* (y) dan variabel *independen* (x) dengan plot PACF
5. Melakukan pembagian data menjadi data *training* dan data *testing*, dengan presentase perbandingan 90%:10%
6. Menerapkan metode SVR untuk memodelkan harga saham dengan mencari *hyperparameter* terbaik melalui optimasi *grid search*. Berikut adalah langkah-langkah yang dilakukan:
 - a. Menentukan fungsi kernel yang akan digunakan, nilai *hyperparameter* kernel, *hyperparameter* cost, serta *hyperparameter* epsilon
 - b. Menentukan nilai *hyperparameter* optimal dengan menggunakan pendekatan *grid-search time series cross validation*.
7. Melakukan prediksi terhadap data *testing*
8. Evaluasi model regresi pada data *testing* menggunakan MAPE
9. Prediksi harga saham 10 hari kedepan
10. Denormalisasi hasil prediksi

4. HASIL DAN PEMBAHASAN

Statistika deskriptif dari data harga penutupan saham harian PT. Adaro Energy Tbk periode 1 Januari 2018 hingga 31 Desember 2022 menunjukkan bahwa sebanyak 1253 data harga penutupan saham harian PT Adaro Energy Tbk akan digunakan pada penelitian ini. Nilai rata-rata harga penutupan saham pada periode 1 Januari 2018 hingga 31 Desember 2022 sebesar Rp 1800,758/lembar saham dan standar deviasinya sebesar Rp 817,4/lembar saham. Harga terendah berada pada Rp 645/lembar saham, sedangkan harga tertinggiya yaitu sebesar Rp 4140/lembar saham, artinya harga saham penutupan ADRO pada periode tersebut berada pada rentang Rp 645/lembar saham - Rp 4140/lembar saham. Adapun grafik pergerakan harga penutupan saham harian PT Adaro Energy Tbk dari 1 Januari 2018 hingga 31 Desember 2022:



Gambar 4. Grafik Data Harga Penutupan Saham ADRO

Grafik pada gambar 4 terlihat bahwa harga penutupan saham harian PT Adaro Energy Tbk sangat fluktuatif sehingga tidak terlihat ada tren ataupun musiman. Grafik tersebut juga memperlihatkan bahwa harga terendah berada pada data ke-574 yaitu pada tanggal 19 Maret 2020. Anjloknya harga saham ADRO merupakan salah satu efek dari pandemi COVID-19. Pada tahun 2020, kondisi ekonomi yang sulit secara umum, termasuk

di industri batu bara, menimbulkan tekanan besar pada permintaan dan harga batu bara secara global. Dampak pandemi, seperti penurunan aktivitas ekonomi, penutupan industri, dan perubahan kebutuhan energi, menyebabkan penurunan permintaan batu bara secara luas. Akibatnya, harga batu bara juga mengalami penurunan di pasar global (Fernando, 2021). Pada gambar 4 juga terlihat bahwa harga saham ADRO perlahan merangkak naik mulai September 2021. Kenaikan ini terjadi karena musim dingin dan krisis batu bara di Cina yang berdampak pada harga batu bara secara global. Dalam laporan kinerja keuangan kuartal ketiga 2021, manajemen Adaro mencatat bahwa permintaan batu bara masih tinggi selama periode tersebut (Sari, 2021). Pada tahun 2022 terus naik hingga menyentuh harga tertinggi yang terlihat berada pada data ke-1193 yaitu pada tanggal 7 Oktober 2022.

Sebelum dilakukan pengolahan data, data akan dinormalisasi terlebih dahulu menjadi rentang antara 0 sampai 1. Normalisasi data yang digunakan yaitu menggunakan *min-max normalization*. Perhitungan berdasarkan persamaan (2) dengan $X_{min} = 645$ dan $X_{max} = 4140$. Contoh perhitungan normalisasi data harga penutupan saham ADRO dengan *min-max normalization*:

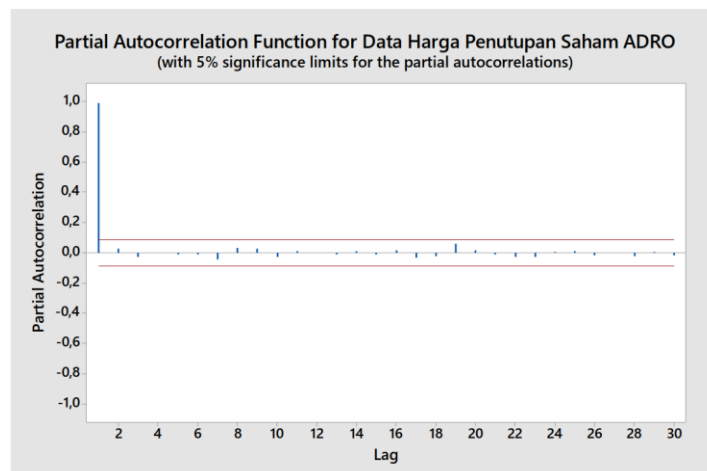
$$X' = \frac{(X - X_{min})}{(X_{max} - X_{min})} = \frac{(X - 645)}{(4140 - 645)}$$

$$X'_1 = \frac{(1860 - 645)}{(4140 - 645)} = 0,347639$$

$$\dots$$

$$X'_{1253} = \frac{(3850 - 645)}{(4140 - 645)} = 0,917024$$

Metode yang akan digunakan dalam penelitian ini menerapkan tipe *learning* yaitu *supervised learning*. Pada proses *supervised learning*, algoritma membutuhkan variabel *input* dan variabel *output* yang telah diketahui untuk dipelajari. Penentuan variabel input dilakukan dengan mengidentifikasi lag yang terputus atau keluar dari garis batas plot PACF (*Partial Autocorrelation Function*). Gambar 5 menampilkan plot PACF untuk data yang sudah dinormalisasi.



Gambar 5. Plot PACF Data Normalisasi Harga Penutupan Saham Harian ADRO

Lag yang signifikan ditandai oleh garis yang berada di luar selang kepercayaan (garis lurus berwarna merah). Gambar 5 memperlihatkan bahwa lag yang signifikan berada pada lag 1. Oleh karena itu, variabel independent (x) yang digunakan pada penelitian ini ialah x_{t-1} dan variabel dependent (y) ialah x_t .

Proporsi data *training* dan data *testing* diperoleh berdasarkan hasil *trial and error* beberapa proporsi yaitu 70%:30%, 80%:20%, dan 90%:10%. Hasil percobaan ketiga proporsi didapatkan model terbaik yang nantinya akan digunakan pada penelitian ini yaitu 90%:10%, sehingga data yang akan digunakan untuk *training* sebanyak 1126 dan jumlah data *testing* yang akan digunakan sebanyak 126. *Hyperparameter* yang akan digunakan pada tahap *loose grid* dapat dilihat pada Tabel 1, sedangkan untuk *hyperparameter* pada tahap *finer grid* masing-masing kernel dapat dilihat pada Tabel 2.

Tabel 1. Nilai *Hyperparameter* Tahap *Loose grid*

Kernel	Hyperparameter	Nilai
Linier	C	$2^{-5}, 2^{-4}, 2^{-3}, 2^{-2}, 2^{-1}, 2^0, 2^1, 2^2, 2^3, 2^4, 2^5$
	ε	$10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}$
Polynomial	C	$2^{-5}, 2^{-4}, 2^{-3}, 2^{-2}, 2^{-1}, 2^0, 2^1, 2^2, 2^3, 2^4, 2^5$
	ε	$10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}$
	d	2, 3, 4
RBF	C	$2^{-5}, 2^{-4}, 2^{-3}, 2^{-2}, 2^{-1}, 2^0, 2^1, 2^2, 2^3, 2^4, 2^5$
	ε	$10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}$
	γ	$2^{-5}, 2^{-4}, 2^{-3}, 2^{-2}, 2^{-1}, 2^0, 2^1, 2^2, 2^3, 2^4, 2^5$
Sigmoid	C	$2^{-5}, 2^{-4}, 2^{-3}, 2^{-2}, 2^{-1}, 2^0, 2^1, 2^2, 2^3, 2^4, 2^5$
	ε	$10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}$
	γ	$2^{-5}, 2^{-4}, 2^{-3}, 2^{-2}, 2^{-1}, 2^0, 2^1, 2^2, 2^3, 2^4, 2^5$

Tabel 2. Nilai *Hyperparameter* Tahap *Finer grid*

Kernel	Hyperparameter	Nilai
Linier	C	$2^1, 2^{1,25}, 2^{1,5}, 2^{1,75}, 2^2, 2^{2,25}, 2^{2,5}, 2^{2,75}, 2^3$
	ε	10^{-4}
Polynomial	C	$2^3, 2^{3,25}, 2^{3,5}, 2^{3,75}, 2^4, 2^{4,25}, 2^{4,5}, 2^{4,75}, 2^5$
	ε	10^{-2}
	d	2
RBF	C	$2^3, 2^{3,25}, 2^{3,5}, 2^{3,75}, 2^4, 2^{4,25}, 2^{4,5}, 2^{4,75}, 2^5$
	ε	10^{-4}
	γ	$2^{-4}, 2^{-3,75}, 2^{-3,5}, 2^{-3,25}, 2^{-3}, 2^{-2,75}, 2^{-2,5}, 2^{-2,25}, 2^{-2}$
Sigmoid	C	$2^4, 2^{4,25}, 2^{4,5}, 2^{4,75}, 2^5, 2^{5,25}, 2^{5,5}, 2^{5,75}, 2^6$
	ε	10^{-4}
	γ	$2^{-5}, 2^{-4,75}, 2^{-4,5}, 2^{-4,25}, 2^{-4}, 2^{-3,75}, 2^{-3,5}, 2^{-3,25}, 2^{-3}$

Pada kernel linier *Hyperparameter* yang digunakan adalah *cost* (C) dan *epsilon* (ε). *Hyperparameter* yang optimal ditentukan melalui optimasi *grid search* yang dipadukan dengan *time series cross validation*. Metode *grid search* yang digunakan pada penelitian ini ada dua tahap yakni *loose grid* dan *finer grid*. Nilai *hyperparameter* yang digunakan dalam tahap *loose grid* kernel linier dapat dilihat pada Tabel 1. Pada tahap *loose grid*, diperoleh nilai optimal dari *hyperparameter* model SVR dengan kernel linier sebesar $C = 4$ atau 2^2 dan $\varepsilon = 0,0001$ atau 10^{-4} . Nilai-nilai ini akan digunakan untuk melanjutkan ke tahap pencarian grid yang lebih halus (*Finer grid*). *Finer grid* adalah fase pencarian *grid* yang dilakukan dengan mempertimbangkan nilai-nilai sekitar C dan ε yang diperoleh pada tahap *loose grid* sebelumnya. Tabel 2 berisi nilai *hyperparameter* yang digunakan pada tahap *finer grid* kernel linier. Melalui proses *grid search* pada tahap ini, menghasilkan nilai *hyperparameter* optimal yang akan digunakan pada model SVR dengan kernel liner yaitu $C = 4$ atau 2^2 dan $\varepsilon = 0,0001$ atau 10^{-4} .

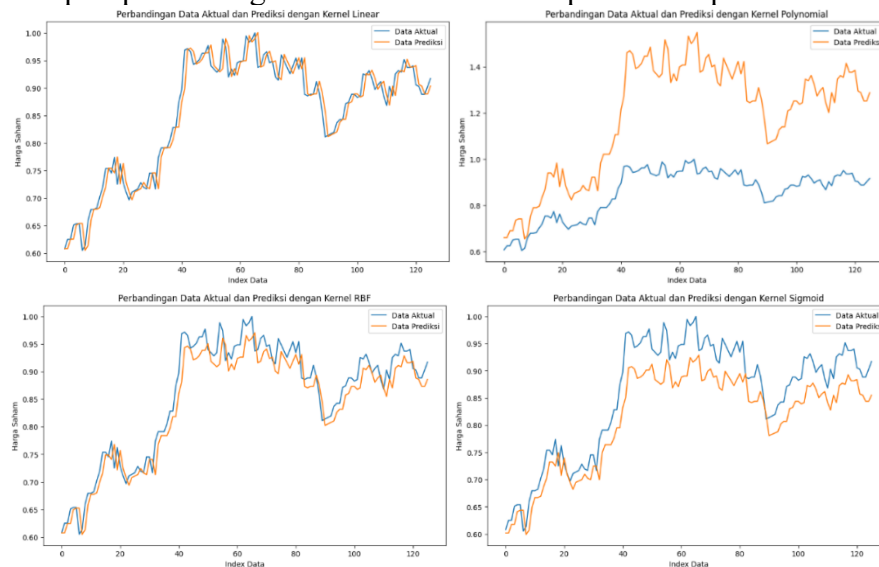
Pada kernel polynomial langkah yang dilakukan sama dengan pada kernel linier, hanya saja ada penambahan *hyperparameter* yang digunakan pada kernel polynomial yaitu *degree* (d). Nilai *hyperparameter* yang digunakan dalam tahap *loose grid* kernel polinomial dapat dilihat pada Tabel 1. Pada tahap *loose grid*, diperoleh nilai optimal dari

hyperparameter model SVR dengan kernel polynomial sebesar $C = 16$ atau 2^4 , $\varepsilon = 0,01$ atau 10^{-2} , dan $d = 2$. Nilai-nilai ini akan digunakan untuk melanjutkan ke tahap pencarian grid yang lebih halus (*Finer grid*). Tabel 2 berisi nilai *hyperparameter* yang digunakan pada tahap *finer grid* kernel polinomial. Melalui proses *grid search* pada tahap ini, menghasilkan nilai *hyperparameter* optimal yang akan digunakan pada model SVR dengan kernel polynomial yaitu $C = 22,62742$ atau $2^{4,5}$, $\varepsilon = 0,01$ atau 10^{-2} , dan $d = 2$.

Pada kernel RBF *Hyperparameter* yang digunakan adalah *cost* (C), *epsilon* (ε), dan *gamma* (γ). Tabel 1 menampilkan nilai *hyperparameter* yang digunakan dalam tahap *loose grid* kernel RBF. Pada tahap *loose grid*, didapatkan nilai optimal dari *hyperparameter* model SVR dengan kernel RBF sebesar $C = 16$ atau 2^4 , $\varepsilon = 0,0001$ atau 10^{-4} , dan $\gamma = 0,125$ atau 2^{-3} . Nilai-nilai ini akan digunakan untuk melanjutkan ke tahap pencarian grid yang lebih halus (*Finer grid*). Tabel 2 berisi nilai *hyperparameter* yang digunakan pada tahap *finer grid* kernel RBF. Melalui proses *grid search* pada tahap ini, menghasilkan nilai *hyperparameter* optimal yang akan digunakan pada model SVR dengan kernel RBF yaitu $C = 8$ atau 2^3 , $\varepsilon = 0,0001$ atau 10^{-4} , dan $\gamma = 0,14865$ atau $2^{-2,75}$.

Pada kernel sigmoid *hyperparameter* yang digunakan sama dengan kernel RBF yaitu *cost* (C), *epsilon* (ε), dan *gamma* (γ). Tabel 1 menampilkan nilai *hyperparameter* yang digunakan dalam tahap *loose grid* kernel sigmoid. Pada tahap *loose grid*, didapatkan nilai optimal dari *hyperparameter* model SVR dengan kernel sigmoid sebesar $C = 32$ atau 2^5 , $\varepsilon = 0,0001$ atau 10^{-4} , dan $\gamma = 0,0625$ atau 2^{-4} . Nilai-nilai ini akan digunakan untuk melanjutkan ke tahap pencarian grid yang lebih halus (*Finer grid*). Tabel 2 berisi nilai *hyperparameter* yang digunakan pada tahap *finer grid* kernel sigmoid. Melalui proses *grid search* pada tahap ini, menghasilkan nilai *hyperparameter* optimal yang akan digunakan pada model SVR dengan kernel sigmoid yaitu $C = 19,02731$ atau $2^{4,25}$, $\varepsilon = 0,0001$ atau 10^{-4} , dan $\gamma = 0,125$ atau 2^{-3} .

Berikut plot perbandingan data aktual dan data prediksi tiap kernel:



Gambar 6. Plot Perbandingan Prediksi dan Aktual Data Testing

Pemilihan model terbaik dengan kernel yang paling optimal dilakukan melalui evaluasi model berdasarkan nilai MAPE.

Tabel 3. Pemilihan Model Terbaik

Kernel	MAPE		
	70:30	80:20	90:10
Linear	2,061 %	2,031 %	1,608 %
Polynomial	27,891 %	41,492%	30,396%
RBF	2,067 %	2,067 %	2,154 %
Sigmoid	2,058 %	2,034 %	4,188 %

Tabel 3 merupakan hasil perhitungan MAPE dari beberapa kernel tiap proporsi. MAPE terkecil sebesar 1,608% terlihat berada pada perbandingan data *training* dan *testing* 90:10 dengan kernel linier, artinya model terbaik yang akan digunakan untuk prediksi adalah model SVR menggunakan kernel linier dengan presentase data *training* 90% dan data *testing* 10%.

Tahap selanjutnya yaitu melakukan peramalan harga penutupan saham harian ADRO menggunakan model terbaik yang sudah dibangun yaitu model SVR kernel linier dengan nilai $\text{cost} = 4$ dan $\text{epsilon} = 0,0001$

Tabel 4. Hasil Prediksi dengan Kernel Linier

Hasil Prediksi	Denormalisasi hasil Prediksi	Aktual
0,90404589	3804,64	3590
0,90471079	3806,964	3520
0,90537638	3809,29	3300
0,90604266	3811,619	3100
0,90670963	3813,95	3140
0,9073773	3816,284	3030
0,90804565	3818,62	3140
0,90871471	3820,958	3180
0,90938445	3823,299	3090
0,9100549	3825,642	3140

5. KESIMPULAN

Model prediksi harga penutupan saham harian ADRO dengan metode SVR telah dibuat berdasarkan *hyperparameter* optimal dari tuning *hyperparameter* menggunakan *grid search*. Empat model yang diperoleh yaitu model dengan kernel linier, polinomial, RBF, dan sigmoid. Pemilihan model terbaik berdasarkan MAPE data *testing* tiap kernel, secara berturut-turut nilainya yaitu 1,608%, 30,396%, 2,154%, dan 4,188%. Model terbaik yang diperoleh adalah SVR dengan kernel linier dengan *hyperparameter* $\text{Cost} = 4$ dan $\epsilon = 0,0001$. Sehingga, kesimpulan yang didapatkan yaitu model SVR dengan kernel linier paling cocok digunakan untuk prediksi harga penutupan saham harian ADRO dibanding dengan kernel yang lainnya karena memiliki nilai MAPE yang paling kecil.

DAFTAR PUSTAKA

- Abe, S. 2010. *Support vector Machines for Pattern Classification*. In Scientific American (Vol. 224, Issue 4).
- Fernando, A. 2021. *Pandemi 2020, Laba Bersih ADRO Jeblok 64%, Gegara Apa?*. Tersedia di: <https://www.cnbcindonesia.com>. (Diakses pada Mei 2023).

- Fitrinanda, G., Djunaidy, A. 2022. *Peramalan Harga Saham PT Adaro Energy Indonesia Tbk yang Mempertimbangkan Faktor Kurs Dolar Amerika Menggunakan Bidirectional Long-Short Term Memory*. Jurnal Teknik Vol. 11, No. 3.
- Hadi, S. R. 2013. *Sukses Membeli Saham Tanpa Modal Secara Otodidak*. Jakarta Timur: Laskar Aksara.
- Heryadi, Y., Wahyono, T. 2020. *Machine Learning: Konsep dan Implementasi*. Yogyakarta: Gava Media
- Lai, L. K. C., Liu, J. N. K. 2010. *Stock forecasting using Support vector Machine*. International Conference on Machine learning and Cybernetics, ICMLC 2010, 4(July), 1607–1614.
- Mahdevari, S., Shahriar, K., Yagiz, S., Shirazi, M. A. 2014. *A Support vector Regression model for predicting tunnel boring machine penetration rates*. International Journal of Rock Mechanics and Mining Sciences, 72, 214–229.
- Muthiah, H. 2022. *Support Vector Regression Analysis for Electricity Load*. 1(November), 70–82.
- Patriya, E. 2020. *Implementasi Support vector Machine Pada Prediksi Harga Saham Gabungan (Ihsg)*. Jurnal Ilmiah Teknologi Dan Rekayasa, 25(1), 24–38.
- Priliani, E. M., Putra, A. T., Muslim, M. A. 2018. *Forecasting Inflation Rate Using Support Vector Regression (SVR) Based Weight Attribute Particle Swarm Optimization (WAPSO)*. Scientific Journal of Informatics, 5(2), 118–127.
- Raschka, S., Mirjalili, V. 2019. *Python Machine Learning: Machine learning and Deep Learning with Python, scikit-learn, and TensorFlow 2*. Packt Publishing.
- Santoso, B. 2013. *Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Yogyakarta: Graha Ilmu
- Sari, I. N. 2021. *Kinerja Adaro 2021 Membara Lewat Lonjakan Harga Batu Bara*. Tersedia di: <https://katadata.co.id>. (Diakses pada Mei 2023).
- Smola, A. J., Scholkopf, B. 2004. *A Tutorial on Support vector Regression*. Australian National University, 10(7), 813–825.
- Vijh, M., Chandola, D., Tikkiwal, V. A., Kumar, A. 2020. *Stock Closing Price Prediction using Machine learning Techniques*. Procedia Computer Science, 167(2019), 599–606.
- Woschnagg, E., Cipan, J. 2004. *Evaluating forecast accuracy*. Forecasting Economic Time Series, 52–78.
- Yahoo Finance. 2023. PT Adaro Energy Indonesia Tbk (ADRO.JK). Tersedia di: <https://finance.yahoo.com>. (Diakses pada Januari 2023).
- Yasin, H., Prahutama, A., Utami, T. W. 2014. *Prediksi Harga Saham Menggunakan Support Vector Regression Dengan Algoritma Grid search*. Media Statistika, 7(1).
- Yudhawan, D. H., Purwaningsih, T. 2020. *Developing Support vector Regression model to forcast stock prices of mining companies in Indonesia*. Jurnal Informatika, 14(2), 43.