

ANALISIS METODE *HYBRID* ARIMA–SVR PADA INDEKS HARGA SAHAM GABUNGAN

Clara Vista Magdalena Sihombing, Shantika Martha, Nur'ainul Miftahul Huda

INTISARI

Data runtun waktu dapat mengandung komponen linier atau nonlinier. Komponen linier dapat dimodelkan dengan Autoregressive Integrated Moving Average (ARIMA), sedangkan komponen nonlinier dimodelkan dengan Support Vector Regression (SVR). Namun, pada kenyataannya, runtun waktu memiliki komponen linier dan nonlinier secara bersamaan. Salah satu metode yang dapat mengatasi komponen linier dan nonlinier adalah Hybrid ARIMA–SVR yang merupakan metode gabungan dari model ARIMA dan SVR. Pada penelitian ini, metode yang digunakan adalah Hybrid ARIMA–SVR dengan studi kasus harga penutupan Indeks Harga Saham Gabungan (IHSG). Rentang waktu IHSG dari 1 Januari 2020 sampai 31 Oktober 2021. IHSG diasumsikan tidak stasioner, serta mengandung komponen linier dan nonlinier secara bersamaan. Tujuan dari penelitian ini adalah memodelkan data IHSG menggunakan model Hybrid ARIMA–SVR, serta membandingkan metode ARIMA dengan metode Hybrid ARIMA–SVR pada data IHSG. Proses penelitian dimulai dengan membentuk model ARIMA dari data training IHSG dan menentukan residual ARIMA. Apabila residual ARIMA nonlinier, maka dapat dibentuk model SVR dari residual ARIMA. Hasil dari penelitian ini yaitu model ARIMA (3,1,3) dan parameter SVR yaitu parameter C , γ , dan ϵ berturut-turut sebesar $2^{-6,75}$, $2^{-2,25}$, dan 0,1. Nilai MAPE model ARIMA untuk data training dan testing sebesar 0,997% dan 0,733%, sedangkan untuk Hybrid ARIMA–SVR sebesar 0,971% dan 0,708%. Kesimpulan yang diperoleh yaitu model terbaik untuk IHSG adalah Hybrid ARIMA–SVR, karena memiliki nilai MAPE yang lebih kecil dibandingkan model ARIMA. Oleh karena itu, nilai IHSG diprediksi lima hari ke depan dengan model Hybrid ARIMA–SVR.

Kata Kunci: model hybrid, linier, nonlinier

PENDAHULUAN

Runtun waktu merupakan serangkaian nilai observasi yang diambil selama kurun waktu tertentu dalam interval yang sama panjang. Analisis runtun waktu merupakan analisis data dengan metode kuantitatif untuk menentukan pola data waktu yang lampau dari suatu variabel. Analisis runtun waktu dapat dilakukan dengan berbagai metode peramalan, salah satunya adalah *Autoregressive Integrated Moving Average* (ARIMA). Metode ARIMA terdiri dari metode *Autoregressive* (AR) dan *Moving Average* (MA) dengan proses diferensiasi [1]. Kekurangan dari metode ARIMA yaitu penurunan keakuratan dalam memprediksi data runtun waktu dengan komponen nonlinier [2]. Selain metode ARIMA, terdapat metode *Support Vector Regression* (SVR) yang dapat memprediksi data runtun waktu nonlinier [3].

Berdasarkan penjelasan metode ARIMA dan SVR, dapat disimpulkan bahwa runtun waktu dapat berpola linier atau nonlinier. Namun pada kenyataannya, runtun waktu dapat berpola linier dan nonlinier secara bersamaan [2]. Oleh karena itu, diperlukan model kombinasi dari model linier dan nonlinier. *Hybrid ARIMA–SVR* adalah salah satu metode peramalan yang menggabungkan model linier dan nonlinier. Model ARIMA sebagai model linier, sedangkan SVR sebagai model nonlinier [4]. Pada penelitian ini, data yang diasumsikan memiliki model linier dan nonlinier secara bersamaan adalah data Indeks Harga Saham Gabungan (IHSG) saat pandemi *Covid-19*.

Terjadinya pandemi *Covid-19* berdampak bagi kehidupan masyarakat, termasuk dalam hal investasi. Saham merupakan salah satu instrumen investasi yang diminati oleh masyarakat. Indeks saham adalah ukuran statistik yang mencerminkan keseluruhan pergerakan harga atas sekumpulan saham. Fluktuasi IHSG mengalami penurunan pada awal pemberitaan kasus pandemi *Covid-19* di Indonesia [5]. Pandemi *Covid-19* mengakibatkan ketidakpastian pada harga saham, oleh sebab itu

diperlukan solusi yang dapat mengurangi risiko kerugian saham. Salah satu solusi yang dapat dilakukan adalah dengan melakukan prediksi IHSG di masa mendatang. Oleh karena itu, penelitian ini bertujuan untuk memodelkan data IHSG menggunakan model *Hybrid* ARIMA–SVR. Selain itu, untuk membandingkan metode ARIMA dengan metode *Hybrid* ARIMA–SVR pada data IHSG.

Studi kasus pada penelitian ini menggunakan data harga penutupan IHSG harian dari 1 Januari 2020 sampai 31 Oktober 2021. Penentuan parameter terbaik model ARIMA menggunakan estimasi parameter yang signifikan dan asumsi residual *white noise*, sedangkan parameter model SVR menggunakan metode *Grid Search* dengan *Kernel Radial Basis Function* (RBF). Proses dari penelitian ini dimulai dengan pembagian data *training–testing* dan uji stasioneritas. Kemudian dibentuk model ARIMA terbaik dan menentukan residual ARIMA. Apabila residual ARIMA nonlinier, maka dapat dibentuk model SVR dari residual ARIMA. Model *Hybrid* ARIMA–SVR diperoleh dari penggabungan model ARIMA dan SVR. Apabila residual *Hybrid* ARIMA–SVR linier, maka model *Hybrid* ARIMA–SVR merupakan model terbaik.

STASIONERITAS DATA

Model stasioner pada data runtun waktu merupakan suatu model yang seluruh sifat-sifat statistiknya tidak berubah dalam pergeseran waktu. Stasioneritas data dibagi menjadi dua, yaitu stasioner dalam varians dan rata-rata. Pengujian data stasioner dalam varians menggunakan transformasi *Box–Cox*. Persamaan transformasi *Box–Cox* sebagai berikut [1]:

$$Z_t = \begin{cases} \ln(Y_t), & \lambda = 0 \\ \frac{Y_t^\lambda - 1}{\lambda}, & \lambda \neq 0 \end{cases}$$

dengan λ adalah parameter transformasi, Y_t adalah data runtun waktu ke- t , dan Z_t adalah hasil transformasi data runtun waktu ke- t . Transformasi *Box–Cox* dapat dijelaskan dengan Tabel 1.

Tabel 1. Transformasi *Box–Cox*

Nilai λ	-1,0	-0,5	0	0,5	1,0
Transformasi	$\frac{1}{Y_t}$	$\frac{1}{\sqrt{Y_t}}$	$\ln(Y_t)$	$\sqrt{Y_t}$	Y_t

Apabila nilai λ bernilai satu, maka data tidak perlu melakukan transformasi. Data runtun waktu dengan λ sebesar satu disebut stasioneritas dalam varians. Cara untuk menguji stasioneritas data dalam rata-rata adalah dengan menguji *Augmented Dickey–Fuller* (ADF) yang merupakan suatu uji akar unit [5]. Apabila data tidak stasioner dalam rata-rata, maka dilakukan proses diferensiasi. Diferensiasi dapat dilakukan hingga beberapa periode sampai data runtun waktu bersifat stasioner. Notasi penting dalam proses diferensiasi adalah operator *backward shift* (B). Notasi B ditunjukkan pada persamaan berikut:

$$\begin{aligned} BY_t &= Y_{t-1} \\ BBY_t &= B^2Y_t = Y_{t-2} \\ B^dY_t &= Y_{t-d} \end{aligned}$$

Data runtun waktu yang telah didiferensiasi pada orde ke-2 ditunjukkan dengan persamaan berikut:

$$Y_t'' = Y_t' - Y_{t-1}' = (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2}) = Y_t - 2BY_t + B^2Y_t = (1 - B)^2Y_t$$

Jika data telah stasioner pada diferensiasi orde ke- d , maka persamaan diferensiasi yang diperoleh adalah sebagai berikut:

$$Y_t^{(d)} = (1 - B)^dY_t$$

dengan $Y_t^{(d)}$ merupakan data runtun waktu yang melalui proses diferensiasi orde ke- d [1].

UJI LINIERITAS

Data runtun waktu terdiri dari komponen linier dan nonlinier, apabila setelah dilakukan pemodelan linier pada data, residual dari model tersebut bersifat nonlinier. Pada penelitian ini, setelah data

dimodelkan dengan ARIMA, dilakukan pengujian linieritas dari residual ARIMA menggunakan uji Terasvirta [6]. Berikut ini adalah hipotesis dari uji Terasvirta:

H_0 : residual tidak mengandung pola nonlinier

H_1 : residual mengandung pola nonlinier

Statistik uji :

$$F_{hitung} = \frac{(SSR_0 - SSR_1)/m}{SSR_1/(n - p - m - 1)}$$

dengan SSR_0 adalah jumlah kuadrat residual, SSR_1 adalah jumlah kuadrat residual dengan pendekatan ekspansi taylor. n merupakan banyaknya data, p merupakan banyaknya variabel prediktor, dan m merupakan prediktor tambahan dari hasil pendekatan ekspansi taylor. Kriteria pengambilan keputusan adalah H_0 ditolak apabila nilai F_{hitung} lebih besar dibanding $F_{(a,m,n-p-m-1)}$, atau nilai p -value kurang dari nilai taraf signifikansi (5%).

MODEL AUTOREGRESSIVE INTEGRATED MOVING AVERAGE (ARIMA)

Model ARIMA merupakan kombinasi dari model AR dengan orde p , MA dengan orde q , dan proses diferensiasi (d). Metode AR menyatakan suatu ramalan sebagai fungsi dari nilai-nilai sebelumnya dan nilai residual saat ini dari runtun waktu tertentu. Persamaan AR(p) adalah sebagai berikut:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \varepsilon_t$$

Metode MA menyatakan suatu ramalan sebagai fungsi dari nilai-nilai residual dari runtun waktu sebelumnya dan saat ini. Persamaan MA(q) adalah sebagai berikut:

$$Y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}$$

Secara umum, model ARIMA (p,d,q) untuk data runtun waktu Y_t dinyatakan pada Persamaan (1).

$$\phi_p(B)(1 - B)^d Y_t = \theta_q(B)\varepsilon_t \tag{1}$$

dengan $\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$ adalah koefisien AR dan $\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$ adalah koefisien MA. $(1 - B)^d$ merupakan diferensiasi orde ke- d , dengan B adalah operator *backward shift*, serta ε_t adalah residual data runtun waktu ke- t [1].

MODEL SUPPORT VECTOR REGRESSION (SVR)

Support Vector Regression (SVR) adalah algoritma yang dapat mengenali pola data runtun waktu yang dapat memprediksi komponen nonlinier dari data runtun waktu dengan fungsi kernel [7]. Ide dasar dari SVR yaitu dengan membagi data menjadi data *training* dan *testing*. Data *training* bertujuan untuk menentukan suatu fungsi dengan batasan deviasi tertentu, sehingga dapat menghasilkan prediksi yang mendekati nilai aktual. Misalkan terdapat m data *training*, (x_i, y_i) dengan $i=1, 2, \dots, m$. Data *input* merupakan x_i dan data *output* adalah y_i berdasarkan x_i yang bersesuaian. Fungsi regresi dari metode SVR untuk komponen nonlinier dituliskan pada Persamaan (2) [8].

$$f(x_i) = \sum_{i=1}^m (\alpha_i - \alpha_i^*) \varphi(x_i) \varphi(x) + b, \tag{2}$$

dengan $f(x_i)$ merupakan fungsi SVR, α_i dan α_i^* merupakan koefisien *lagrange*, $\varphi(x_i)$ merupakan fungsi yang memetakan x_i , $\varphi(x)$ merupakan fungsi yang memetakan x , serta b adalah bias. x_i merupakan data *input* ke- i , sedangkan x merupakan data *input* untuk memetakan fungsi ke ruang fitur [8].

Fungsi regresi $f(x_i)$ diasumsikan dapat menempatkan seluruh titik (x_i, y_i) dengan akurasi ε . Pada umumnya $\varphi(x_i)$ pada Persamaan (2) tidak diketahui. Oleh karena itu, $\varphi(x_i)$ dapat diatasi dengan kernel yang ditunjukkan pada Persamaan (3) [9].

$$K(x_i, x) = \varphi(x_i) \cdot \varphi(x) \tag{3}$$

Metode kernel adalah perkalian skalar dari $\varphi(x_i)$ dan $\varphi(x)$ dalam ruang fitur. Kernel RBF merupakan fungsi kernel yang digunakan ketika terdapat komponen nonlinier pada data. Fungsi kernel RBF dijelaskan pada Persamaan (4).

$$K(x_i, x) = \exp(-\gamma \|x_i - x\|^2), \text{ dengan } \gamma > 0 \quad (4)$$

dengan $K(x_i, x)$ merupakan kernel RBF dan γ merupakan parameter *gamma*. Apabila fungsi kernel RBF pada Persamaan (3) dan (4) disubstitusikan ke Persamaan (2), maka menghasilkan fungsi regresi SVR pada Persamaan (5).

$$f(x_i) = \sum_{i=1}^m (\alpha_i - \alpha_i^*) \exp(-\gamma \|x_i - x\|^2) + b \quad (5)$$

dengan $0 < \alpha_i \leq C, 0 < \alpha_i^* \leq C$

SVR dengan fungsi kernel RBF memiliki tiga parameter yaitu parameter *cost* (C), *gamma* (γ), dan *epsilon* (ε). Metode yang digunakan untuk mencari parameter SVR yang optimal adalah metode *grid search*. *Grid search* merupakan metode untuk menentukan parameter optimal untuk model SVR. Metode *grid search* melalui dua tahap pengerjaan, yaitu tahap *loose grid* dan *finer grid* [10]. *Loose grid* adalah proses pemilihan nilai C dan γ dengan pangkat bilangan bulat. Pemilihan parameter C dan γ menggunakan urutan angka eksponensial. Misalnya rentang parameter C adalah $2^{-5}, 2^{-3}, \dots, 2^{15}$, sedangkan rentang parameter γ adalah $2^{-15}, 2^{-13}, \dots, 2^5$ [11]. Penentuan parameter ε menggunakan rentang dari 0,01 sampai 0,1 yang merupakan bilangan *real* [2]. *Finer grid* merupakan proses tahapan lanjutan dari *loose grid* dengan pemilihan nilai C dan γ yaitu nilai pangkat berada pada rentang parameter optimum dari tahap *loose grid*. Parameter ε dari tahap *loose grid* tetap digunakan pada tahap *finer grid*. Pemilihan parameter optimum berdasarkan pada nilai *error* terendah.

MODEL HYBRID ARIMA–SVR

Model *Hybrid ARIMA–SVR* merupakan kombinasi model ARIMA dan model SVR [4]. Model *Hybrid ARIMA–SVR* dapat digunakan apabila diasumsikan model data runtun waktu terdiri dari komponen linier dan nonlinier. Model *Hybrid ARIMA–SVR* dinyatakan dalam persamaan berikut:

$$Y_t = L_t + N_t$$

dengan Y_t adalah data runtun waktu ke- t , L_t merupakan komponen linier ke- t (ARIMA) dan N_t merupakan komponen nonlinier ke- t (SVR). Setelah menentukan model ARIMA, kemudian diperoleh residual ARIMA yang diasumsikan mengandung komponen nonlinier. Persamaan residual ARIMA sebagai berikut:

$$\varepsilon_t = Y_t - \hat{L}_t$$

dengan ε_t merupakan residual ARIMA dan \hat{L}_t merupakan nilai estimasi dari model ARIMA dalam waktu ke- t . Nilai residual yang memuat komponen nonlinier dimodelkan dengan SVR. Ketika model ARIMA dan SVR telah diperoleh, maka hasil prediksi dengan model *Hybrid ARIMA–SVR* ditunjukkan dalam Persamaan (6).

$$\hat{Y}_t = \hat{L}_t + \hat{N}_t \quad (6)$$

dengan \hat{N}_t nilai estimasi dari model SVR dalam waktu ke- t dan \hat{Y}_t adalah nilai estimasi dari model *Hybrid ARIMA–SVR* [2].

AKURASI METODE PERAMALAN

Akurasi metode peramalan bertujuan untuk mengevaluasi hasil dari prediksi yang telah dilakukan. Penelitian ini menggunakan *Mean Absolute Percentage Error* (MAPE) sebagai akurasi dari metode peramalan. Berikut persamaan untuk menentukan nilai MAPE [2]:

$$MAPE = \frac{\sum_{t=1}^n \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right|}{n} \times 100\%$$

dengan \hat{Y}_t adalah hasil prediksi data runtun waktu ke- t dan n adalah banyaknya data [2]. Berdasarkan penilaian MAPE, kesimpulan akurasi metode peramalan dijelaskan pada Tabel 2.

Tabel 2. Kriteria MAPE

Persentase MAPE	$\leq 10\%$	11% - 20%	21% - 50%	$> 50\%$
Tingkat Akurasi	sangat akurat	Akurat	kurang akurat	tidak akurat

STUDI KASUS

Data yang digunakan pada penelitian ini merupakan data sekunder, yaitu data harga penutupan Indeks Harga Saham Gabungan (IHSG) dari *website yahoo finance* [12]. Rentang waktu IHSG dari 1 Januari 2020 sampai 31 Oktober 2021. Data IHSG dinyatakan dalam variabel Y_t .



Gambar 1. Plot Data IHSG

Berdasarkan Gambar 1, terjadi penurunan nilai IHSG pada bulan Maret 2020 akibat pandemi *Covid-19* di Indonesia. Penurunan IHSG terjadi karena penurunan harga saham di seluruh sektor industri (ditunjukkan dengan gambar lingkaran). IHSG terlihat mulai stabil setelah diterapkannya kebijakan *new normal* pada Juni 2020.

Tabel 3. Statistika Deskriptif IHSG

Statistik Deskriptif	Rata-rata	Nilai Maksimum	Nilai Minimum
Nilai (Rp)	5658,446	6658,768	3937,632

Berdasarkan Tabel 3, rata-rata dari IHSG sebesar Rp5658,446. Nilai IHSG terendah sebesar Rp3937,632 terjadi pada tanggal 24 Maret 2020. Hal ini disebabkan adanya pandemi *Covid-19* di Indonesia yang menyebabkan penurunan harga saham. Nilai IHSG tertinggi terjadi pada tanggal 18 Oktober 2021 sebesar Rp6658,768. Nilai IHSG perlahan-lahan kembali stabil pada tahun 2021. Hal ini diakibatkan kebijakan *new normal* serta diadakannya vaksin *Covid-19* oleh pemerintah.

PEMBENTUKAN MODEL ARIMA

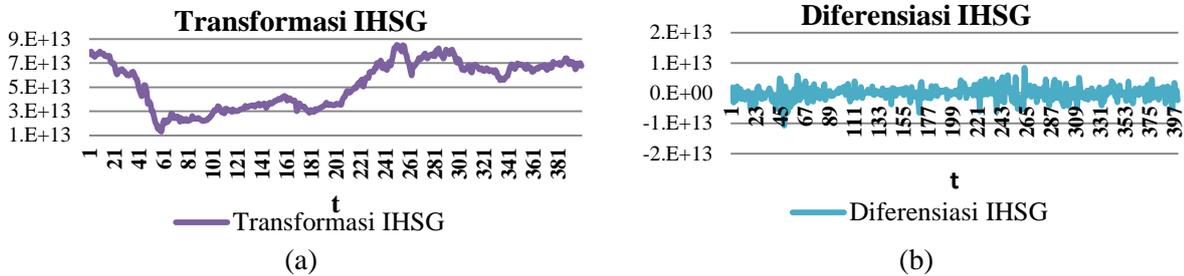
Langkah awal untuk membentuk model ARIMA yaitu menguji kestasioneritas data. Uji stasioner dalam varians dapat dilakukan dengan cara mencari nilai lambda (λ) data *training* IHSG. Nilai λ yang diperoleh sebesar 3,81, sehingga perlu dilakukan transformasi. Data yang telah ditransformasi disebut Z_t . Setelah data telah stasioner dalam varians, maka dapat dilanjutkan pengujian kestasioneran data dalam rata-rata. Uji stasioner dalam rata-rata dapat dilakukan dengan uji *Augmented Dickey-Fuller* (ADF) yang dijelaskan pada Tabel 4.

Tabel 4. Uji ADF Data *Training* IHSG

Statistik Uji	Nilai	
	Sebelum Diferensiasi	Sesudah Diferensiasi
Dickey-Fuller	-3,095	-6,109
<i>p-value</i>	0,115	0,010

Tabel 4 menunjukkan nilai *p-value* pada data *training* IHSG sebelum diferensiasi lebih besar dibandingkan nilai α yaitu 0,05, sehingga H_0 tidak ditolak. Kesimpulannya adalah data tidak stasioner

dalam rata-rata. Apabila data tidak stasioner dalam rata-rata, maka dilakukan diferensiasi untuk menstasionerkan data. Setelah melakukan diferensiasi, nilai *p-value* pada uji ADF lebih kecil dibandingkan nilai α , sehingga H_0 ditolak. Kesimpulannya adalah data *training* IHSB telah stasioner dalam rata-rata. Plot data setelah ditransformasi dan diferensiasi ditunjukkan pada Gambar 2.



Gambar 2. (a) Plot Transformasi IHSB dan (b) Diferensiasi IHSB

Gambar 2(a) merupakan fluktuasi dari transformasi data *training* IHSB, sedangkan Gambar 2(b) merupakan grafik diferensiasi dari hasil transformasi data *training* IHSB. Berdasarkan Gambar 2(b) terlihat bahwa data telah berfluktuasi di sekitar suatu nilai rata-rata, memiliki varians yang konstan, serta tidak bergantung pada waktu.

Identifikasi model ARIMA didasarkan pada plot korelogram ACF dan PACF dari data IHSB yang telah diferensiasi yang ditunjukkan dengan Gambar 3.



Gambar 3. (a) Korelogram ACF dan (b) PACF

Gambar 3(a) menunjukkan korelogram ACF dari data IHSB yang dapat dilihat bahwa ACF terputus setelah *lag* ke-3. Korelogram PACF pada Gambar 3(b) menunjukkan bahwa PACF terputus setelah *lag* ke-3. Oleh karena itu, beberapa model ARIMA yang diperoleh dari korelogram ACF dan PACF adalah ARIMA (0,1,3), ARIMA (3,1,0), dan ARIMA (3,1,3). Estimasi parameter model ARIMA dinyatakan signifikan apabila seluruh parameter memiliki nilai $Pr(>|z|)$ kurang dari nilai taraf signifikan (0,050). Parameter model ARIMA dijelaskan pada Tabel 5.

Tabel 5. Estimasi parameter Model ARIMA

No	Model	Parameter	Estimasi	$Pr(> z)$	Kesimpulan
1	ARIMA (0,1,3)	θ_1	0,093	0,063	Tidak Signifikan
		θ_2	-0,127	0,011	Signifikan
		θ_3	0,127	0,006	Signifikan
2	ARIMA (3,1,0)	ϕ_1	0,079	0,113	Tidak Signifikan
		ϕ_2	-0,100	0,043	Signifikan
		ϕ_3	0,120	0,016	Signifikan
3	ARIMA (3,1,3)	ϕ_1	-0,613	$5,43 \times 10^{-5}$	Signifikan
		ϕ_2	0,754	$2,01 \times 10^{-10}$	Signifikan
		ϕ_3	0,741	$< 2,20 \times 10^{-16}$	Signifikan
		θ_1	0,738	$2,61 \times 10^{-5}$	Signifikan
		θ_2	-0,7780	$8,70 \times 10^{-13}$	Signifikan
		θ_3	0,7758	$6,70 \times 10^{-10}$	Signifikan

Tabel 5 menunjukkan seluruh parameter ARIMA (3,1,3) signifikan. Oleh karena itu, residual ARIMA (3,1,3) dilakukan uji diagnostik untuk mengidentifikasi sifat *white noise* pada residual. *White noise* terpenuhi apabila residual berdistribusi normal dengan rata-rata nol dan varians konstan, serta tidak berkorelasi. Uji diagnostik residual ARIMA (3,1,3) ditunjukkan pada Gambar 4.



Gambar 4. Uji Diagnostik Residual, (a) ACF Residual dan (b) Uji Normalitas Residual ARIMA

Gambar 4 menunjukkan uji diagnostik residual ARIMA. Gambar 4(a) menunjukkan nilai ACF dari residual ARIMA (3,1,3). Seluruh *lag* pada ACF dan PACF terletak di dalam garis signifikansi, sehingga dapat disimpulkan residual ARIMA (3,1,3) tidak saling berkorelasi. Gambar 4(b) menunjukkan plot uji normalitas dari residual ARIMA. Pada Gambar 4(b) terlihat bahwa residual ARIMA berdistribusi normal, karena sebaran data mendekati garis diagonal. Oleh karena itu, ARIMA (3,1,3) memenuhi asumsi *white noise*.

Berdasarkan estimasi parameter dan uji diagnostik residual, maka ARIMA (3,1,3) merupakan model ARIMA terbaik. Persamaan model ARIMA (3,1,3) adalah:

$$\begin{aligned} \hat{L}_t &= \phi_1(Y_{t-1} - Y_{t-2}) + \phi_2(Y_{t-2} - Y_{t-3}) + \phi_3(Y_{t-3} - Y_{t-3}) + Y_{t-1} - \theta_1(\varepsilon_{t-1}) - \theta_2(\varepsilon_{t-2}) - \theta_3(\varepsilon_{t-3}) \\ &= -0,6132(Y_{t-1}) + 0,6132(Y_{t-2}) + 0,7542(Y_{t-2}) - 0,7542(Y_{t-3}) + 0,7414(Y_{t-3}) - 0,7414(Y_{t-4}) + Y_{t-1} - 0,7385(\varepsilon_{t-1}) + 0,7780(\varepsilon_{t-2}) + 0,7758(\varepsilon_{t-3}). \end{aligned}$$

Kemudian *invers* nilai transformasi dengan λ sebesar 3,81. Apabila nilai $Y_t = \sqrt[3,81]{3,81(Z_t) + 1}$, Persamaan dari model ARIMA (3,1,3) ditunjukkan pada Persamaan (8).

$$\begin{aligned} \hat{L}_t &= -0,6132(\sqrt[3,81]{3,81(Z_{t-1}) + 1}) + 0,6132(\sqrt[3,81]{3,81(Z_{t-2}) + 1}) + 0,7542(\sqrt[3,81]{3,81(Z_{t-2}) + 1}) - 0,7542(\sqrt[3,81]{3,81(Z_{t-3}) + 1}) + 0,7414(\sqrt[3,81]{3,81(Z_{t-3}) + 1}) - 0,7414(\sqrt[3,81]{3,81(Z_{t-4}) + 1}) + \sqrt[3,81]{3,81(Z_{t-1}) + 1} - 0,7385(\varepsilon_{t-1}) + 0,7780(\varepsilon_{t-2}) + 0,7758(\varepsilon_{t-3}). \end{aligned} \tag{8}$$

dengan persamaan residual model ARIMA adalah $\varepsilon_t = Y_t - \hat{L}_t$. Z_t merupakan data transformasi, Y_t merupakan data asli, dan \hat{L}_t merupakan model ARIMA.

PEMBENTUKAN MODEL SVR

Residual ARIMA diubah ke dalam bentuk *time lag*. Pembagian *time lag* dibagi menjadi satu sampai enam *lag*. Pemilihan *time lag* terbaik menggunakan *trial and error* dengan akurasi RMSE [2]. Berdasarkan nilai RMSE, *lag* empat memiliki nilai RMSE terkecil, sehingga nilai *lag* empat digunakan sebagai data *input* pada model SVR. Setelah menentukan *time lag* residual, maka dilanjutkan dengan uji linieritas residual dengan uji Terasvirta. Hasil dari uji Terasvirta dijelaskan pada Tabel 6.

Tabel 6. Uji Terasvirta Residual ARIMA

Statistik Uji	Nilai
<i>F</i>	3,443
<i>p-value</i>	$1,499 \times 10^{-8}$

Nilai *p-value* pada uji Terasvirta sebesar $2,34 \times 10^{-9}$ yang lebih kecil dari nilai taraf signifikansi sebesar 0,05. Kesimpulannya adalah residual ARIMA mengandung pola nonlinier. Oleh karena itu, dapat dilanjutkan dengan metode SVR.

Parameter SVR yaitu *cost* (*C*), *gamma* (γ), dan *epsilon* (ϵ) ditentukan dengan metode *grid search*. Pada tahap *loose grid*, rentang parameter *cost* dan *gamma* yaitu $2^{-9}, 2^{-7}, \dots, 2^7, 2^9$, sedangkan parameter *epsilon* yaitu 0,01; 0,02; \dots ; 0,1. Penentuan parameter optimal menggunakan nilai residual terkecil. Nilai residual kombinasi parameter SVR dapat dilihat pada Tabel 7.

Tabel 7. Nilai Residual Parameter SVR pada Tahap *Loose Grid*

		$\epsilon = 0,1$						
$C \backslash \gamma$		2^{-9}	2^{-7}	2^{-5}	2^{-3}	2^{-1}	\dots	2^9
2^{-9}		5136,52	5136,90	5138,30	5145,12	5170,25	\dots	5693,21
2^{-7}		5136,79	5137,80	5142,64	5161,16	5230,18	\dots	6106,96
2^{-5}		5136,72	5138,47	5144,90	5193,82	5269,91	\dots	10081,61
2^{-3}		5136,33	5131,25	5138,77	5182,94	5272,72	\dots	8733,07
\vdots		\vdots	\vdots	\vdots	\vdots	\vdots	\ddots	\vdots
2^9		5136,02	5135,53	5132,01	5133,37	5140,84	\dots	5138,28

Berdasarkan Tabel 7, parameter *cost* (*C*), *gamma* (γ), dan *epsilon* (ϵ) berturut-turut sebesar $2^{-7}, 2^{-3}$, dan 0,1, dengan residual sebesar 5131,25. Pada tahap *finer grid*, rentang parameter *cost* (*C*) yang digunakan yaitu $2^{-8}, 2^{-7,75}, 2^{-7,5}, \dots, 2^{-6,25}, 2^{-6}$, sedangkan parameter *gamma* (γ) yaitu $2^{-4}, 2^{-3,75}, 2^{-3,5}, \dots, 2^{-2,25}, 2^{-2}$. Parameter *epsilon* (ϵ) tetap bernilai 0,1. Tahap *finer grid* ditunjukkan pada Tabel 8.

Tabel 8. Nilai Residual Parameter SVR pada Tahap *Finer Grid*

$C \backslash \gamma$	2^{-8}	\dots	$2^{-7,75}$	$2^{-6,75}$	$2^{-6,5}$	$2^{-6,25}$	2^{-6}
2^{-4}	5137,32	\dots	5136,65	5136,03	5136,92	5135,84	5135,63
\vdots	\vdots	\ddots	\vdots	\vdots	\vdots	\ddots	\vdots
$2^{-2,25}$	5132,06	\dots	5131,15	5128,24	5128,56	5129,12	5129,96
2^{-2}	5132,20	\dots	5131,32	5128,80	5129,35	5130,31	5131,41

Tabel 8 menunjukkan parameter *cost* (*C*) sebesar $2^{-6,75}$, parameter *gamma* (γ) sebesar $2^{-2,25}$, dan parameter *epsilon* (ϵ) sebesar 0,1. Nilai residual yang diperoleh sebesar 5128,24. Oleh karena itu, diperoleh model SVR terbaik dengan nilai parameter *C*, γ , dan ϵ berturut-turut sebesar $2^{-6,75}, 2^{-2,25}$, dan 0,1. Model SVR ditunjukkan pada Persamaan (9).

$$\hat{N}_t = \sum_{t=5}^{396} (\alpha_t - \alpha_t^*) \exp(-2^{-2,25} \|x_t - x\|^2) + b, \tag{9}$$

dengan $0 < \alpha_t \leq 2^{-6,75}, 0 < \alpha_t^* \leq 2^{-6,75}$.

HYBRID ARIMA-SVR

Setelah menentukan parameter terbaik dari model ARIMA dan SVR, selanjutnya adalah menggabungkan kedua model tersebut. Berdasarkan Persamaan (8) dan (9), maka *Hybrid ARIMA-SVR* ditunjukkan pada Persamaan (10).

$$\begin{aligned} \hat{Y}_t &= \hat{L}_t + \hat{N}_t \\ &= -0,6132 \left(\sqrt[3,81]{3,81(Z_{t-1}) + 1} \right) + 0,6132 \left(\sqrt[3,81]{3,81(Z_{t-2}) + 1} \right) + \\ &\quad 0,7542 \left(\sqrt[3,81]{3,81(Z_{t-2}) + 1} \right) - 0,7542 \left(\sqrt[3,81]{3,81(Z_{t-3}) + 1} \right) + \\ &\quad 0,7414 \left(\sqrt[3,81]{3,81(Z_{t-3}) + 1} \right) - 0,7414 \left(\sqrt[3,81]{3,81(Z_{t-4}) + 1} \right) + \\ &\quad \sqrt[3,81]{3,81(Z_{t-1}) + 1} - 0,7385(\epsilon_{t-1}) + 0,7780(\epsilon_{t-2}) + \\ &\quad 0,7758(\epsilon_{t-3}) + \sum_{t=5}^{396} (\alpha_t - \alpha_t^*) \exp(-2^{-2,25} \|x_t - x\|^2), \end{aligned} \tag{10}$$

dengan $0 < \alpha_t \leq 2^{-6,75}, 0 < \alpha_t^* \leq 2^{-6,75}$.

Interpretasi Persamaan (10) adalah komponen linier pada data *training* IHSG yang dimodelkan dengan ARIMA. Nilai IHSG waktu ke- t dipengaruhi oleh nilai IHSG satu sampai empat hari sebelumnya dan residual nilai IHSG satu sampai tiga hari sebelumnya. Komponen nonlinier dimodelkan oleh SVR. Nilai IHSG waktu ke- t dipengaruhi nilai residual ARIMA ke- t , dengan kernel RBF sebesar $(\gamma) 2^{-2,25}$, dan koefisien *lagrange* diantara 0 sampai $2^{-6,75}$.

Setelah memperoleh model *Hybrid* ARIMA–SVR, dilakukan kembali pembagian *time lag* dan pengujian linieritas residual dari *Hybrid* ARIMA–SVR. Uji terasvirta yang dijelaskan pada Tabel 9.

Tabel 9. Uji Terasvirta Residual *Hybrid* ARIMA–SVR

Statistik Uji	Nilai
F	1,212
p -value	0,298

Berdasarkan Tabel 9, kesimpulan yang diperoleh yaitu residual dari *Hybrid* ARIMA–SVR telah linier, sehingga model *Hybrid* ARIMA–SVR tersebut merupakan model terbaik.

PERBANDINGAN MODEL ARIMA DAN *HYBRID* ARIMA–SVR

Perbandingan model ARIMA dan *Hybrid* ARIMA–SVR menggunakan nilai MAPE yang dijelaskan pada Tabel 10.

Tabel 10. Nilai MAPE Model ARIMA dan *Hybrid* ARIMA–SVR

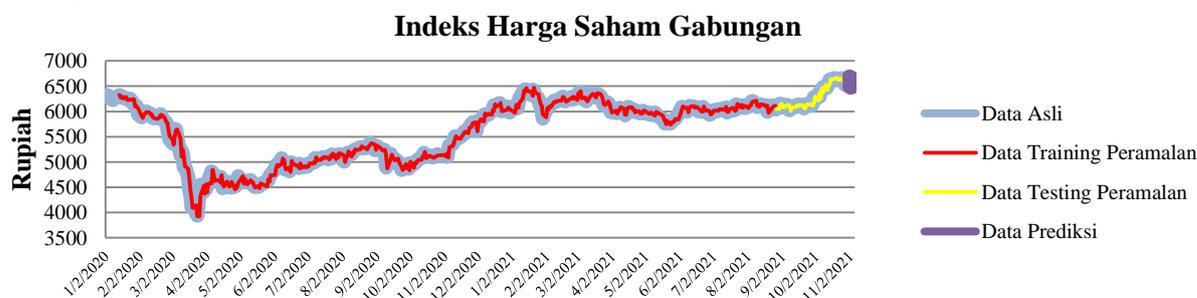
Model	Data Training	Data Testing
ARIMA	0,997%	0,733%
<i>Hybrid</i> ARIMA–SVR	0,971%	0,708%

Tabel 10 menunjukkan nilai MAPE pada model ARIMA dan *Hybrid* ARIMA–SVR. Berdasarkan Tabel 2, kedua model dapat menghasilkan nilai prediksi IHSG dengan sangat akurat. Akan tetapi, nilai MAPE model *Hybrid* ARIMA–SVR lebih kecil dibandingkan model ARIMA, sehingga model *Hybrid* ARIMA–SVR merupakan model terbaik. Prediksi IHSG dengan metode *Hybrid* ARIMA–SVR ditunjukkan pada Tabel 11.

Tabel 11. Prediksi IHSG dengan Model *Hybrid* ARIMA–SVR

Periode	Tanggal	Prediksi IHSG (Rp)
1	1 November 2021	6688,690
2	2 November 2021	6488,568
3	3 November 2021	6649,979
4	4 November 2021	6517,412
5	5 November 2021	6571,904

Grafik prediksi dengan *Hybrid* ARIMA–SVR ditampilkan pada Gambar 5.



Gambar 5. Grafik Hasil Prediksi *Hybrid* ARIMA–SVR

Gambar 5 menunjukkan fluktuasi IHSG apabila diprediksi dengan model *Hybrid* ARIMA–SVR. Data *training* dan *testing* berturut-turut ditunjukkan dengan warna merah dan kuning, sedangkan data

prediksi ditunjukkan dengan warna ungu. Pola dari data *training* dan *testing* IHSG mengikuti pola data asli dengan baik.

PENUTUP

Metode *Hybrid* ARIMA–SVR merupakan metode gabungan dari ARIMA dan SVR, dengan metode ARIMA bertujuan untuk mengatasi komponen linier dan metode SVR yang mengatasi komponen nonlinier. Studi kasus yang digunakan pada penelitian adalah IHSG dengan rentang Januari 2020 sampai Oktober 2021. Kesimpulan yang diperoleh yaitu model *Hybrid* ARIMA–SVR, dengan ARIMA (3,1,3) dan parameter optimal SVR yaitu parameter C sebesar $2^{-6,75}$, parameter γ sebesar $2^{-2,25}$, dan parameter ε sebesar 0,01. Model *Hybrid* ARIMA–SVR merupakan model terbaik, karena memiliki nilai MAPE yang lebih baik dibandingkan model ARIMA. Nilai MAPE yang diperoleh dari model *Hybrid* ARIMA–SVR untuk data *training* dan data *testing* berturut-turut sebesar 0,971% dan 0,708%.

DAFTAR PUSTAKA

- [1]. Wei WWS. *Time Series Analysis: Univariate and Multivariate Methods*. Second Edition. New York: Addison Wesley; 2006.
- [2]. Purnama DI. Prediksi Harga Emas Saat Pandemi Covid-19 menggunakan Model Hybrid Autoregressive Integrated Moving Average–Support Vector Regression. *Jambura Journal of Mathematics*. 2021; 3(1):52–65.
- [3]. Fitria IM, Darwis S. Pemodelan menggunakan Autoregressive Integrated Moving Average (ARIMA) dan Support Vector Regression (SVR) pada Vibrasi Bearing. *Prosiding Statistika*. 2020; 6(2):49–56.
- [4]. Zhang Y, Luo L, Yang J, Dunhu L, Kong R, Feng Y. A Hybrid ARIMA–SVR Approach for Forecasting Emergency Patient Flow. *Journal of Ambient Intelligence and Humanized Computing*. 2018; 10(1):3315–3323.
- [5]. Rusdi. Uji Akar-Akar Unit dalam Model Runtun Waktu Autoregresif. *Statistika*. 2011; 11(2):67–78.
- [6]. Prabowo H, Suhartono, Prastyo, DD. The Performance of Ramsey Test, White Test, and Terasvirta Test in Detecting Nonlinearity. *Inferensi*. 2020; 3(1):2721–3862.
- [7]. Yasin H, Prahutama A, Utami TW. Prediksi Harga Saham menggunakan Support Vector Regression dengan Algoritma Grid Search. *Media Statistika*. 2014; 7(1):29–35.
- [8]. Sriyana, Martha S, Sulistyanyingsih E. Prediksi Nilai Tukar Dolar Amerika Serikat Terhadap Rupiah Dengan Metode Support Vector Regression (SVR). *Bimaster*. 2019; 8(1): 1–10.
- [9]. Vapnik VN. *The Nature of Statistical Learning Theory*. New York: Springer; 1995.
- [10]. Purnama DI, Hendarsin OP. Prediksi Jumlah Penumpang Berangkat melalui Transportasi Udara di Sulawesi Tengah menggunakan Support Vector Regression (SVR). *Jambura Journal of Mathematics*. 2020; 2(2):49–59.
- [11]. Hsu CH, Chang CC, Lin CJ. *A Practical Guide to Support Vector Classification*. Taipei: Department of Computer Science National Taiwan University; 2016.
- [12]. Yahoo Finance. Histori Harga Penutupan Indeks Harga Saham Gabungan [Internet]. 2021 [cited 2021 Nov 2]. Available from: [finace.yahoo.com/quote/%5EJKSE/history/](https://finance.yahoo.com/quote/%5EJKSE/history/).

CLARA VISTA MAGDALENA SIHOMBING : Jurusan Matematika FMIPA Untan, Pontianak
 claraavistaa23@student.untan.ac.id
 SHANTIKA MARTHA : Jurusan Matematika FMIPA Untan, Pontianak
 shantika.martha@math.untan.ac.id
 NUR'AINUL MIFTAHUL HUDA : Jurusan Matematika FMIPA Untan, Pontianak
 nur'ainul@fmipa.untan.ac.id
