

Peramalan Harga Emas Saat Pandemi Covid-19 Menggunakan Model *Hybrid Autoregressive Integrated Moving Average - Support Vector Regression*

Drajat Indra Purnama^{1*}

¹ Badan Pusat Statistik (BPS) Kabupaten Parigi Moutong,
Jl. Pakabata Komplek Perkantoran Bambalemo, Parigi, Parigi Moutong, Sulawesi Tengah, Indonesia

* Penulis Korespondensi. Email: drajatindrapurnama@bps.go.id

ABSTRAK

Investasi emas merupakan salah satu investasi yang menjadi favorit dimasa pandemi Covid 19 seperti sekarang ini. Hal ini dikarenakan harga emas yang nilainya relatif fluktuatif tetapi menunjukkan tren peningkatan. Investor dituntut pandai dalam berinvestasi emas, mampu memprediksi peluang dimasa yang akan datang. Salah satu model peramalan data deret waktu adalah model *Autoregressive Integrated Moving Average* (ARIMA). Model ARIMA baik digunakan pada data yang berpola linear tetapi jika digunakan pada data data nonlinear keakuratannya menurun. Untuk mengatasi permasalahan data nonlinear dapat menggunakan model *Support Vector Regression* (SVR). Pengujian linearitas pada data harga emas menunjukkan adanya pola data linear dan nonlinear sekaligus sehingga digunakan kombinasi ARIMA dan SVR yaitu model *hybrid* ARIMA-SVR. Hasil peramalan menggunakan model *hybrid* ARIMA-SVR menunjukkan hasil lebih baik dibanding model ARIMA. Hal ini dibuktikan dengan nilai MAPE model *hybrid* ARIMA-SVR lebih kecil dibandingkan nilai MAPE model ARIMA. Nilai MAPE model *hybrid* ARIMA-SVR sebesar 0,355 pada data *training* dan 4,001 pada data *testing*, sedangkan nilai MAPE model ARIMA sebesar 0,903 pada data *training* dan 4,076 pada data *testing*.

Kata Kunci:

ARIMA, SVR, Hybrid ARIMA-SVR, Grid Search, Harga Emas

ABSTRACT

Gold investment is one of the favorite investments during the Covid 19 pandemic as it is today. This is because the price of gold is relatively volatile but shows an increasing trend. Investors are required to be smart in investing in gold, able to predict future opportunities. One of the time series data forecasting models is the *Autoregressive Integrated Moving Average* (ARIMA) model. The ARIMA model is good for use on linear patterned data but if it is used on nonlinear data the accuracy decreases. To solve the problem of nonlinear data, you can use the *Support Vector Regression* (SVR) model. The linearity test on the gold price data shows that there are linear and nonlinear data patterns at the same time so that a combination of ARIMA and SVR is used, namely the ARIMA-SVR hybrid model. Forecasting results using the ARIMA-SVR hybrid model show better results than the ARIMA model. This is evidenced by the MAPE value of the ARIMA-SVR hybrid model which is smaller than the MAPE value of the ARIMA model. The MAPE value of the ARIMA-SVR hybrid model is 0.355 on the training data and 4.001 on the testing data, while the MAPE value of the ARIMA model is 0.903

in the training data and 4.076 in the testing data.

Keywords:

ARIMA, SVR, Hybrid ARIMA-SVR, Grid Search, Gold Price

Format Sitasi:

D.I. Purnama, "Peramalan Harga Emas Saat Pandemi Covid-19 Menggunakan Model Hybrid Autoregressive Integrated Moving Average - Support Vector Regression," *Jambura J. Math.*, vol. 3, no. 1, pp. 52-65, 2021

1. Pendahuluan

Investasi menurut Kamus Besar Bahasa Indonesia (KKBI) berarti penanaman uang atau modal dalam suatu perusahaan atau proyek untuk tujuan memperoleh keuntungan. Akan tetapi makna investasi di zaman sekarang tidak melulu berhubungan dengan penanaman modal atau uang. Ada sedikit pergeseran makna dalam hal investasi. Investasi bisa diartikan bahwa seseorang membeli barang yang memiliki nilai dan berharap barang tersebut dapat mempertahankan nilai atau bahkan meningkat nilainya dikemudian hari. Bagi sebagian orang yg "berduit" membeli tas mewah, mobil mewah atau bahkan rumah mewah bisa dianggap sebuah investasi untuk masa depan dengan harapan harga barang tersebut akan naik di masa depan. Selain itu, terdapat investasi lain yg lazim dilakukan oleh masyarakat yakni dengan membeli logam mulia atau emas.

Emas saat ini tetap menjadi favorit investasi yang menjanjikan. Hal ini disebabkan emas adalah barang bernilai yang cenderung meningkat nilainya setiap tahun, bahkan jarang nilainya menurun sehingga investasi emas dalam waktu yang cukup lama sangat menjanjikan dibandingkan investasi lainnya. Menurut Habibah [1], nilai investasi emas yang tidak terpengaruh inflasi, harga emas memiliki tren meningkat, transaksinya fleksibel, menjadikan emas alternatif berinvestasi.

Seiring dengan munculnya pandemi virus corona atau Covid-19 yang melanda Indonesia investasi emas tetap menjadi pilihan. Ketika perekonomian di Indonesia masih tidak menentu dikarenakan pandemi, harga emas menunjukkan kestabilan harga dan cenderung menguntungkan untuk berinvestasi. Harga emas di awal pandemi di Bulan Maret 2020 berkisar antara 726 ribu rupiah sampai 860 ribu rupiah per gram. Selanjutnya harga emas mengalami kenaikan dan penurunan hingga pada akhir November 2020 mencapai 808 ribu rupiah per gram. Pergerakan data harga emas selama pandemi COVID-19 fluktuatif naik turun sehingga membentuk data deret waktu yg tidak stasioner dan besaran kenaikan yang bervariasi membuat data harga emas juga nonlinear. Diperlukan suatu model yang bisa meramalkan harga emas dimasa depan sehingga dapat menjadi acuan bagi orang dalam berinvestasi emas.

Salah satu model peramalan data deret waktu adalah model *Autoregressive Integrated Moving Average* (ARIMA) yang pertama kali diperkenalkan Box dan Jenkins [2]. Keunggulan ARIMA adalah ketepatannya dalam peramalan jangka pendek dan untuk data runtun waktu non stasioner. Akan tetapi, ARIMA memiliki kekurangan untuk sebagian besar permasalahan yang bersifat nonlinear karena cenderung mengalami penurunan tingkat keakuratannya [3].

Model *Support Vector Regression* (SVR) adalah modifikasi *Support Vector Machine* (SVM) yang dipakai untuk pendekatan regresi [4]. SVR dapat digunakan juga untuk melakukan prediksi pada data deret waktu yang berpola nonlinear. SVR memiliki

keunggulan pada kemampuan mengatasi permasalahan data nonlinear menggunakan trik *kernel*. Terdapat parameter SVR yaitu γ (γ), $cost$ (C) dan ϵ (ϵ) yang nilai optimalnya ditentukan dengan metode *grid search*. *Grid search* adalah kombinasi dari parameter yang diujikan pada suatu model SVR untuk mencari nilai *error* dalam klasifikasi [5].

Permasalahan data deret waktu dalam kehidupan nyata terkadang tidak hanya linear atau nonlinear saja. Terkadang mengandung keduanya (linear dan nonlinear) secara sekaligus. Oleh karena itu, banyak penelitian yang telah menggunakan model *hybrid* yang merupakan kombinasi model linear dengan model nonlinear. Salah satu model *hybrid* adalah model *hybrid* ARIMA-SVR yang merupakan kombinasi model ARIMA dan model SVR. Model *hybrid* ini membentuk komponen linear menggunakan model ARIMA dan membentuk komponen nonlinear menggunakan model SVR sehingga kombinasi komponen linear dan komponen nonlinear membentuk model *hybrid* ARIMA-SVR. Penelitian ini menunjukkan bagaimana membentuk model *hybrid* ARIMA-SVR yang merupakan kombinasi dari model ARIMA dan model SVR. Model *hybrid* ARIMA-SVR yang terbentuk diharapkan mempunyai akurasi peramalan yang lebih baik daripada model ARIMA dalam meramalkan harga emas dimasa pandemik Covid 19.

2. Metode

2.1. Stasioneritas Data

Pengertian data stasioner adalah data deret waktu yang tidak memiliki tren, tidak terdapat pola musiman dan nilai rata-rata serta varians konstan dari waktu ke waktu [6]. Data yang stasioner merupakan data yang memiliki varians tidak terlalu besar serta memiliki kecenderungan nilainya mendekati nilai rata-rata. Pengujian stasioneritas pada rata-rata data deret waktu salah satunya menggunakan uji *Augmented Dickey Fuller* (ADF), dengan alur pengujiannya adalah

1) Hipotesis

$H_0 : \gamma = 1$ (Data tidak stasioner dalam rata-rata)

$H_1 : \gamma < 1$ (Data stasioner dalam rata-rata)

2) Statistik uji

$$T = \frac{\hat{\gamma} - 1}{se(\hat{\gamma})} \sim t_{(n-p-1)}$$

dengan $se(\hat{\gamma})$ adalah standar *error* dari $\hat{\gamma}$ dan p adalah banyaknya parameter.

3) Kriteria uji

Tolak H_0 jika nilai $T > t_{tabel}$, artinya data stasioner dalam rata-rata.

2.2. Uji Linearitas

Pengujian linearitas pada data deret waktu salah satunya menggunakan uji *Ramsey RESET* yang diperkenalkan pertama kali oleh Ramsey [7]. Apabila terdapat variabel bebas X_t dan variabel prediktor Y_t , prosedur pada uji *Ramsey RESET* sebagai berikut

1). Melakukan regresi Y_t pada X_t maka diperoleh model linear

$$Y_t = f_t + e_t, \text{ dengan } f_t = X_t \theta_t \tag{1}$$

- 2). Melakukan regresi e_t pada f_t^k dengan f_t^k merupakan transformasi nonlinear dari f_t , sehingga diperoleh

$$e_t = a_2 f_t^2 + \dots + a_k f_t^k + v_t, \text{ untuk } k \geq 2 \quad (2)$$

Selanjutnya persamaan (2) dimasukkan ke persamaan (1) sehingga diperoleh model alternative

$$Y_t = X_t \theta_t + a_2 f_t^2 + \dots + a_k f_t^k + v_t, \text{ untuk } k \geq 2 \quad (3)$$

Langkah-langkah dalam pengujian linearitas pada uji *Ramsey RESET* adalah

- 1). Hipotesis

$H_0 : a_2 = \dots = a_k = 0$ (Data deret waktu tidak mengandung pola nonlinear)

$H_1 : a_k \neq 0$ (Data deret waktu mengandung pola nonlinear)

- 2). Statistik uji

$$RESET = \frac{[(\mathbf{e}_t^T \mathbf{e}_t - \mathbf{v}_t^T \mathbf{v}_t)/(k - 1)]}{[(\mathbf{v}_t^T \mathbf{v}_t)/(n - k)]}$$

dengan

$\mathbf{e}_t = (e_1, \dots, e_n)$: vektor residual dari model linear pada persamaan (1)

$\mathbf{v}_t = (v_1, \dots, v_n)$: vektor residual dari model alternatif pada persamaan (3)

n : banyaknya pengamatan

k : banyaknya variabel bebas dan prediktor

- 3). Kriteria uji

Tolak H_0 jika $RESET > F_{(k-1, n-k)}$, hal ini berarti data deret waktu yang digunakan mengandung pola nonlinear.

3.3. Model Autoregressive Integrated Moving Average (ARIMA)

Model *Autoregressive Integrated Moving Average* (ARIMA) merupakan model univariat yang biasa digunakan dalam peramalan pada data deret waktu [2]. Model ARIMA adalah merupakan gabungan model *Autoregressive* (AR) dan model *Moving Average* (MA) atau biasa disebut model *Autoregressive Moving Average* (ARMA) pada data yang telah dilakukan *differencing*. Prosedur ARIMA adalah dengan melakukan *differencing* pada data deret waktu nonstasioner sehingga diperoleh data deret waktu stasioner, untuk selanjutnya dilakukan proses ARMA pada data stasioner hasil *differencing* tersebut.

Secara umum, model ARIMA(p, d, q) dengan p menunjukkan orde AR, d adalah orde proses *differencing* dan q menunjukkan orde MA [8], dituliskan sebagai berikut

$$\phi_p(B)(1 - B)^d Z_t = \theta_q(B)\varepsilon_t$$

dengan

ϕ_p : parameter AR

θ_q : parameter MA

B : operator *Backshift*

$(1 - B)^d$: *differencing* orde d

ε_t : residual yang menyebar normal dan independen $\varepsilon_t \stackrel{iid}{\sim} N(0, \sigma^2)$

3.4. Model Support Vector Regression (SVR)

SVR memiliki ide dasar dengan membagi set data menjadi data *training* dan *testing*. Selanjutnya dari data *training* ditentukan suatu fungsi regresi menggunakan batasan deviasi tertentu agar dapat menghasilkan prediksi mendekati nilai sesungguhnya. Konsep SVR didasarkan pada *structural risk minimization* (SRM) yaitu mengestimasi suatu fungsi resiko atau *risk function* yaitu dengan meminimalkan batas atas dari *generalization error*.

Andaikan terdapat sebanyak n set data *training*, (\mathbf{x}_i, y_i) dengan $i = 1, 2, \dots, n$. Sedangkan $\mathbf{x}_i = \{x_1, x_2, \dots, x_p\}^T \in R^n$ merupakan vektor pada *input space* dan $y_i = \{y_1, y_2, \dots, y_n\} \in R$ merupakan nilai *output* pada \mathbf{x}_i yang bersesuaian. Bentuk umum fungsi regresi sempurna dari metode SVR linear dituliskan sebagai berikut

$$f(\mathbf{x}_i) = \mathbf{w} \cdot \mathbf{x}_i + b$$

dengan \mathbf{w} adalah vektor bobot dan b adalah bias.

Menurut Haykin [9], koefisien \mathbf{w} dan b diestimasi dengan meminimalkan fungsi resiko berikut

$$R(f(\mathbf{x}_i)) = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n E_\varepsilon(y_i - f(\mathbf{x}_i))$$

atau dituliskan

$$\min_{\mathbf{w}} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n E_\varepsilon(y_i - f(\mathbf{x}_i)) \tag{4}$$

dengan $\|\mathbf{w}\|$ merupakan reguralisasi atau fungsi yang diminimumkan agar membuat fungsi setipis (*flat*) mungkin. Konstanta C (*Cost*) > 0 merupakan nilai tawar antara ketipisan fungsi f dengan batas atas deviasi lebih dari ε yang bisa ditoleransi [10]. Sedangkan E_ε merupakan ε -insensitive loss function yang dituliskankan sebagai

$$E_\varepsilon(y_i - f(\mathbf{x}_i)) = \begin{cases} |y_i - f(\mathbf{x}_i)| - \varepsilon, & \text{untuk } |y_i - f(\mathbf{x}_i)| \geq \varepsilon \\ 0, & \text{lainnya.} \end{cases}$$

dengan kendala

$$\begin{aligned} y_i &\leq f(\mathbf{x}_i) + \varepsilon \\ y_i &\geq f(\mathbf{x}_i) - \varepsilon \end{aligned}$$

Fungsi regresi f diasumsikan adalah fungsi yang dapat mengaproksimasi semua titik (\mathbf{x}_i, y_i) dengan presisi ε . Semua titik yang terletak pada rentang $f \pm \varepsilon$ disebut *feasible*. Titik yang terletak diluar rentang $f \pm \varepsilon$ disebut *infeasible*. Pada titik *infeasible* ini ditambah dengan variabel *slack* positif ξ_i, ξ_i^* agar dapat mengatasi permasalahan pembatas yang tidak layak dalam masalah optimasi. Selanjutnya persamaan (4) ditranformasikan menjadi bentuk

$$\min_{\mathbf{w}, \xi_i, \xi_i^*} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \tag{5}$$

dengan kendala

$$\begin{aligned} y_i &\leq f(\mathbf{x}_i) + \varepsilon + \xi_i \\ y_i &\geq f(\mathbf{x}_i) - \varepsilon - \xi_i^* \\ \xi_i, \xi_i^* &\geq 0 \end{aligned}$$

Menurut Smola dan Scholkopf [10], pada persamaan (5) memperlihatkan bahwa penentuan parameter \mathbf{w} dan b adalah masalah optimasi *quadratic programming* dengan *lagrangian* sebagai berikut

$$\begin{aligned} L_p = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) - \sum_{i=1}^n (\eta_i \xi_i + \eta_i^* \xi_i^*) - \sum_{i=1}^n \alpha_i (\varepsilon + \xi_i + f(\mathbf{x}_i) - y_i) \\ - \sum_{i=1}^n \alpha_i^* (\varepsilon + \xi_i^* - f(\mathbf{x}_i) + y_i) \end{aligned}$$

dengan kendala dengan $\alpha_i, \alpha_i^*, \eta_i, \eta_i^* \geq 0$ adalah koefisien *lagrange*. *Lagrangian* menghasilkan solusi optimal dari parameter \mathbf{w} berbentuk koefisien *lagrange* α_i dan α_i^* sebagai berikut

$$\mathbf{w} = \sum_{i=1}^n (\alpha_i - \alpha_i^*) \mathbf{x}_i$$

sehingga fungsi regresi metode SVR dapat dituliskan sebagai

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

Selanjutnya *lagrangian* menghasilkan persamaan dual untuk problem optimisasi dari SVR sebagai berikut

$$L_d(\alpha_i, \alpha_i^*) = \sum_{i=1}^n y_i (\alpha_i - \alpha_i^*) - \varepsilon \sum_{i=1}^n (\alpha_i + \alpha_i^*) - \frac{1}{2} \sum_{i,j=1}^n (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) \mathbf{x}_i \mathbf{x}_j$$

dengan $\mathbf{x}_i \mathbf{x}_j$ merupakan perkalian skalar (*dot-product*) dua data dalam data *training*. Menurut Smola dan Scholkopf [10], solusi optimal bias (b) dapat ditentukan dengan kondisi KKT (Karush-Kuhn-Tucker). Untuk estimasi b diperoleh menggunakan kondisi KKT sebagai berikut

$$\begin{aligned} \alpha_i (\varepsilon + \xi_i - y_i + f(\mathbf{x}_i)) &= 0 \\ \alpha_i^* (\varepsilon + \xi_i^* + y_i - f(\mathbf{x}_i)) &= 0 \\ (C - \alpha_i) \xi_i &= 0 \\ (C - \alpha_i^*) \xi_i^* &= 0 \end{aligned}$$

Berdasarkan kondisi KKT diperoleh nilai estimasi akhir dari b adalah $b = y_i - \mathbf{w} \cdot \mathbf{x}_i - \varepsilon$ untuk $0 \leq \alpha_i \leq C$ dan $b = y_i - \mathbf{w} \cdot \mathbf{x}_i + \varepsilon$ untuk $0 \leq \alpha_i^* \leq C$.

Kenyataannya data pada dunia nyata jarang terpisah secara linear atau bersifat *linear separable*, banyak yang bersifat nonlinear. Penggunaan SVR pada kasus nonlinear

adalah memberi pendekatan alternatif dengan melakukan pemetaan data \mathbf{x} pada *input space* ke *feature space* yang berdimensi lebih tinggi menggunakan suatu fungsi φ atau dituliskan $\varphi: \mathbf{x} \mapsto \varphi(\mathbf{x})$ [11]. Fungsi regresi SVR nonlinear dapat dituliskan sebagai

$$f(\mathbf{x}_i) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) \varphi(\mathbf{x}_i) \cdot \varphi(\mathbf{x}) + b$$

Akan tetapi, kesulitan dalam pemetaannya adalah bahwa transformasi φ pada kenyataannya sulit dipahami dan umumnya tidak diketahui. Menurut Vapnik [12], solusi dari permasalahan ini adalah *kernel trick*. *Kernel trick* merupakan perkalian skalar (*dot product*) $\varphi(\mathbf{x}_i) \cdot \varphi(\mathbf{x})$ pada *feature space* dapat digantikan oleh fungsi kernel sebagai berikut

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

Sehingga fungsi regresi SVR dapat dituliskan menjadi

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

Parameter SVR dan fungsi kernel pada model SVR harus diatur dengan tepat karena bisa mempengaruhi akurasi regresi. Kernel yang digunakan pada model SVR penelitian ini adalah kernel *Radial Basis Function* (RBF) yaitu

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

Sehingga fungsi regresi SVR nonlinear menggunakan kernel RBF dituliskan sebagai

$$f(\mathbf{x}_i) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}\|^2) + b$$

dengan $\gamma > 0$. Terdapat tiga parameter SVR dengan fungsi kernel RBF yang harus ditentukan nilainya yaitu γ (*gamma*), ε (*epsilon*) dan C (*cost*). Pada penelitian ini, metode yang digunakan untuk melakukan estimasi parameter optimal adalah metode *grid search*.

3.5. Metode Grid Search

Metode *grid search* secara umum merupakan kombinasi dari parameter yang diujikan pada suatu model SVR untuk mencari nilai *error* dalam klasifikasi [5]. Tujuan dari metode ini adalah mengidentifikasi parameter optimal pada data *training*, sehingga model tersebut dapat memprediksi data *testing* secara akurat. Salah satu pendekatan pada metode *grid search* yaitu menggunakan *cross validation*. Prosedur *cross validation* adalah membagi secara acak data *training* menjadi n subset yang berukuran sama. Sebanyak $n-1$ subset digunakan sebagai data *training* dan sisanya satu subset digunakan sebagai data *testing* [13]. Pasangan parameter dengan akurasi terbaik yang diperoleh dari uji *cross validation* merupakan parameter yang optimal.

Proses untuk melakukan *grid search* secara lengkap memerlukan waktu yang cukup lama sehingga Hsu *et al.* [5], menyarankan untuk melakukan *grid search* dengan dua tahap yaitu *loose grid* dan *finer grid*. *Loose grid* adalah tahapan dimana pemilihan nilai C dan γ dengan pangkat bilangan bulat, sedangkan *finer grid* adalah tahapan selanjutnya dari *loose grid* dimana saat didapat nilai C dan γ dengan *error* terendah, maka *finer grid* menggunakan persekitaran nilai C dan γ tersebut.

3.6. Model Hybrid ARIMA-SVR

Model *hybrid* ARIMA-SVR merupakan kombinasi model ARIMA dan model SVR. Data deret waktu pada model *hybrid* ini, diasumsikan terdiri dari komponen linear dan komponen linear. Menurut Zhang *et al.* [14], secara umum model data deret waktu yang merupakan kombinasi komponen linear dan komponen nonlinear secara sistematis dituliskan sabagai

$$Z_t = L_t + N_t$$

dengan Z_t merupakan data deret waktu pada waktu ke- t , L_t merupakan komponen linear waktu ke- t dan N_t merupakan komponen nonlinear waktu ke- t .

Tahap pertama dalam membentuk model *hybrid* ARIMA-SVR adalah membentuk komponen linear. Komponen linear dimodelkan menggunakan model ARIMA dengan input datanya merupakan input data deret waktu asli. Hasil estimasi dari model ARIMA selanjutnya disebut sebagai komponen linear. Kemudian residual dari model ARIMA diasumsikan mengandung hubungan nonlinear. Residual dari model komponen linear menggunakan model ARIMA dituliskan sebagai

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

dengan ε_t merupakan residual model ARIMA waktu ke- t dan \hat{L}_t merupakan komponen linear waktu ke- t hasil ramalan dari model ARIMA.

Setelah komponen linear terbentuk, langkah selanjutnya adalah membentuk komponen nonlinear. Komponen nonlinear dimodelkan menggunakan model SVR dengan input data merupakan residual dari model ARIMA. Untuk menentukan banyaknya *lag* untuk input data pada SVR didasarkan pada penelitian Balkin [15]. Menurut Balkin [15], banyaknya *lag* untuk input SVR untuk data deret waktu harian adalah 1 *lag* sampai 6 *lag*. Prosedurnya adalah dengan melakukan *trial and error* pada jumlah 1 *lag* sampai 6 *lag*, dipilih jumlah *lag* yang memiliki *error* terkecil. Model SVR dari residual ARIMA dituliskan sebagai

$$\varepsilon_t = f(\varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_{t-n}) + r_t$$

dengan n merupakan jumlah *lag* berpengaruh, f merupakan fungsi nonlinear yang ditentukan oleh SVR dan r_t merupakan residual model SVR pada waktu ke- t . Selanjutnya ε_t menjadi komponen nonlinear waktu ke- t . Sehingga model *hybrid* ARIMA-SVR dapat dinyatakan sebagai

$$\hat{Z}_t = \hat{L}_t + \hat{N}_t$$

dengan \hat{Z}_t merupakan model peramalan *hybrid* ARIMA-SVR, \hat{L}_t merupakan komponen linear waktu ke- t hasil ramalan dari model ARIMA dan \hat{N}_t komponen nonlinear waktu ke- t hasil peramalan $\hat{\varepsilon}_t$.

3. Hasil dan Pembahasan

Data pada penelitian ini adalah data sekunder dari situs www.harga-emas.org yang merupakan data harga emas per gram harian periode 1 Maret 2020 sampai dengan 30 November 2020. Akan tetapi data harga emas pada hari minggu tidak diikutsertakan karena harga emas pada hari minggu sama dengan harga emas pada hari sabtu. Terdapat sebanyak 210 data deret waktu. Data harga emas pada periode 1 Maret 2020 sampai 2 November 2020 (185 Data) digunakan sebagai data *training* sedangkan data pada periode 3 November 2020 sampai 30 November 2020 (25 Data) digunakan sebagai data *testing*. Pengolahan data pada penelitian ini menggunakan program R 3.6.1.

3.1. Stasioneritas Data

Sebelum melakukan analisis data deret waktu syarat yang harus dipenuhi adalah kestasioneran data. Penelitian ini menggunakan uji *Augmented Dickey Fuller* (ADF) untuk uji stasioneritas pada data harga emas. Hasil uji ADF pada data harga emas disajikan pada Tabel 1 berikut ini

Tabel 1. Pengujian ADF data harga emas sebelum dan sesudah *differencing*

Data	Sebelum <i>differencing</i>		Sesudah <i>differencing</i> pertama	
	<i>p-value</i>	Kesimpulan	<i>p-value</i>	Kesimpulan
Harga emas	0,59	Data Tidak Stasioner	0,01	Data Stasioner

Berdasarkan Tabel 1, hasil uji ADF pada data harga emas menunjukkan nilai *p-value* lebih besar dari $\alpha = 0,05$, sehingga H_0 diterima sehingga dapat disimpulkan data harga emas tidak stasioner. Sehingga untuk membuat data stasioner adalah dengan melakukan *differencing*. Setelah *differencing* pertama, hasil uji ADF pada data harga emas hasil *differencing* pertama memiliki *p-value* lebih kecil dari $\alpha = 0,05$, sehingga H_0 ditolak sehingga dapat disimpulkan data harga emas sudah stasioner.

3.2. Pengujian Linearitas

Pengujian linearitas pada uji *Ramsey RESET* pada data harga emas selama periode 1 Maret 2020 sampai dengan 2 November 2020 diperoleh hasil yang ditunjukkan pada Tabel 2.

Tabel 2. Uji *Ramsey RESET* pada data harga emas

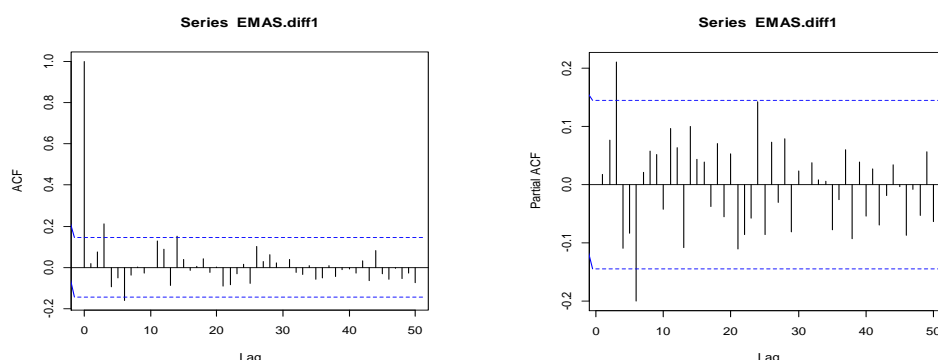
Data	Statistik uji (<i>RESET</i>)	<i>p-value</i>
Harga emas	10,50	0,001418

Berdasarkan Tabel 2 dapat dilihat bahwa nilai statistik uji *Ramsey RESET* lebih besar dari $F_{(0,05;1;182)} = 3,89$ atau nilai *p-value* lebih kecil dari $\alpha = 0,05$ maka dapat disimpulkan H_0 ditolak yang berarti bahwa data harga emas mengandung pola nonlinear.

3.3. Peramalan Harga Emas Menggunakan Model ARIMA

Identifikasi model deret waktu dilakukan untuk menentukan orde waktu untuk model ARIMA yang akan dibangun. Berdasarkan pengujian stasioneritas data harga emas merupakan data yang tidak stasioner sehingga perlu dilakukan *differencing* pertama.

Grafik ACF dan PACF data harga emas setelah dilakukan *differencing* pertama dapat dilihat pada Gambar 1.



Gambar 1. Grafik ACF dan PACF data harga emas *differencing* pertama

Berdasarkan Gambar 1 terlihat bahwa ACF maupun PACF signifikan pada *lag* 3 sehingga ditetapkan model sementara adalah ARIMA (3,1,3). Tahap selanjutnya setelah diperoleh model ARIMA sementara adalah melakukan estimasi parameter model ARIMA yang mungkin untuk menggambarkan sifat data. Menurut prinsip *parsimony* dari pemodelan (model yang baik merupakan model yang memiliki parameter yang sedikit). sedikit) dilakukan pengujian beberapa alternatif model ARIMA yang mungkin. Berdasarkan beberapa kemungkinan model ARIMA yang diestimasi parameternya terdapat tiga model yang memiliki hasil uji parsial semua parameternya signifikan ($|t - \text{hitung}| > t_{0,025;184} = 1,97$) yaitu model ARIMA (1,1,1), ARIMA (1,1,2) dan ARIMA (2,1,2). Hasil uji parsial model ARIMA disajikan pada Tabel 3.

Tabel 3. Hasil uji parsial model ARIMA

No	Model	Parameter	Estimasi Parameter	Standar Error	t-hitung	Kesimpulan
1	ARIMA(1,1,1)	AR(1)	0,65	0,27	2,37	Signifikan
		MA(1)	-0,58	0,29	-2,03	Signifikan
2	ARIMA(1,1,2)	AR(1)	0,47	0,15	3,05	Signifikan
		MA(1)	-0,51	0,15	-3,39	Signifikan
		MA(2)	0,24	0,10	2,27	Signifikan
3	ARIMA (2,1,2)	AR(1)	0,78	0,18	4,37	Signifikan
		AR(2)	-0,50	0,16	-3,09	Signifikan
		MA(1)	-0,82	0,14	-5,87	Signifikan
		MA(2)	0,72	0,13	5,50	Signifikan
		AR(1)	-0,06	0,21	-0,26	Tidak Signifikan
4	ARIMA(3,1,3)	AR(2)	-0,01	0,17	-0,06	Tidak Signifikan
		AR(3)	-0,18	0,18	-1,03	Tidak Signifikan
		MA(1)	0,07	0,19	0,37	Tidak Signifikan
		MA(2)	0,12	0,15	0,80	Tidak Signifikan
		MA(3)	0,50	0,15	3,35	Signifikan

Pengujian diagnostik model merupakan pemeriksaan apakah asumsi dasar model deret waktu sudah terpenuhi. Penelitian ini melakukan uji asumsi residual *white noise*

dengan uji *Portmanteau*. Pada Tabel 4 dapat dilihat hasil pengujian residual *white noise* dengan uji *Portmanteau* pada model ARIMA.

Tabel 4. Pengujian diagnostik model ARIMA

No	Model	Uji <i>white noise</i>	MAPE
1	ARIMA (1,1,1)	Tidak <i>white noise</i>	0,9195947
2	ARIMA (1,1,2)	Tidak <i>white noise</i>	0,9174574
3	ARIMA (2,1,2)	<i>White noise</i>	0,9025762

Berdasarkan pengujian diagnostik model ARIMA diperoleh satu model yang memenuhi asumsi *white noise* yaitu ARIMA (2,1,2) dengan nilai MAPE terkecil sebesar 0,9025762. Sehingga dapat disimpulkan model ARIMA terbaik yang diperoleh adalah ARIMA (2,1,2). Model ARIMA (2,1,2) ini digunakan untuk melakukan peramalan harga emas. Hasil peramalan harga emas yang diperoleh dari model ARIMA (2,1,2) merupakan komponen linear dari model *hybrid* ARIMA-SVR. Setelah diperoleh model ARIMA terbaik dan komponen linear, langkah selanjutnya adalah melakukan perhitungan residual model ARIMA (2,1,2).

3.4. Peramalan Residual ARIMA Terbaik Menggunakan Model SVR

Residual dari model ARIMA (2,1,2) digunakan untuk input data pada model SVR sehingga terbentuk model *hybrid* ARIMA-SVR. Residual ARIMA (2,1,2) tersebut diasumsikan berisi hubungan nonlinear sehingga dimodelkan menggunakan SVR dengan fungsi kernel. Sebelum masuk ke pemodelan menggunakan SVR, langkah pertama yang dilakukan adalah mengkonversikan residual model ARIMA terbaik ke dalam bentuk *time lag* yang nantinya sebagai data input dalam model SVR.

Penelitian ini menggunakan data deret waktu harian sehingga penentuan banyak *time lag* adalah dengan melakukan percobaan (*trial and error*) yaitu mencoba segala kemungkinan banyaknya *time lag* (antara 1 *lag* sampai 6 *lag*) dan memilih jumlah *time lag* yang memberikan nilai *error* terkecil. Proses *trial and error* dilakukan sehingga didapat nilai eror terkecil adalah pada input dua *lag*. Hal ini berarti bahwa untuk memodelkan residual ARIMA (2,1,2) menggunakan SVR adalah dengan input data residual sebanyak dua *lag*.

Pada penelitian ini dibatasi hanya menggunakan fungsi kernel *Radial Basic Function* (RBF). Pada kernel RBF terdapat parameter C , γ dan ε yang harus ditentukan. Untuk mendapatkan nilai parameter yang optimal maka pada penelitian ini digunakan metode *grid search* dengan dua tahapan yaitu *loose grid* dan *finer grid*.

Tahapan awal metode *grid search* adalah menentukan rentang nilai parameter. Pada tahapan *loose grid* digunakan nilai parameter C dan γ dengan pangkat bilangan bulat. Sedangkan nilai ε ditentukan terlebih dahulu. Rentang nilai parameter yang digunakan untuk tahapan *loose grid* pada penelitian ini ditunjukkan pada Tabel 5.

Tabel 5. Rentang nilai parameter tahapan *loose grid* metode *grid search*

Parameter	Rentang Nilai
C	$2^{-9}, 2^{-7}, \dots, 2^3$
γ	$2^{-3}, 2^{-1}, \dots, 2^9$
ε	0,01; 0,02; ...; 0,1

Rentang nilai parameter pada Tabel 5 akan digunakan untuk proses penentuan parameter model SVR. Melalui proses *grid search* pada tahapan *loose grid* diperoleh nilai optimal dari parameter model SVR adalah $C = 2^{-1}$, $\gamma = 2^9$ dan $\varepsilon = 0,02$. Setelah diperoleh parameter optimal model SVR pada tahapan *loose grid*, maka langkah selanjutnya dilakukan *grid search* menggunakan tahapan *finer grid*. *Finer grid* merupakan tahapan *grid search* menggunakan persekitaran nilai C , γ dan ε yang diperoleh pada tahapan *loose grid*.

Tabel 6. Rentang nilai parameter tahapan *finer grid* metode *grid search*

Parameter	Rentang nilai
C	$2^{-1,75}, 2^{-1,5}, 2^{-1,25}, 2^{-1}, 2^{-0,75}, 2^{-0,5}, 2^{-0,25}$
γ	$2^{8,25}, 2^{8,5}, 2^{8,75}, 2^9$
ε	0,02

Rentang nilai parameter yang digunakan pada tahapan *finer grid* ditunjukkan pada Tabel 6. Melalui proses *grid search* pada tahapan *finer grid* diperoleh nilai optimal dari parameter model SVR adalah $C = 2^{-0,5}$, $\gamma = 2^{8,25}$ dan $\varepsilon = 0,02$. Nilai parameter yang diperoleh menggunakan *grid search* pada tahapan *finer grid* merupakan parameter optimal pada model SVR. Sehingga model SVR terbaik yang diperoleh adalah model SVR menggunakan kernel *Radial Basic Function* (RBF) dengan parameternya adalah $C = 2^{-0,5}$, $\gamma = 2^{8,25}$ dan $\varepsilon = 0,02$.

Analisis yang dilakukan selanjutnya adalah melakukan peramalan residual model ARIMA (2,1,2) menggunakan model SVR terbaik. Hasil peramalan residual model ARIMA (2,1,2) menggunakan model SVR terbaik disebut sebagai komponen nonlinear dari model *hybrid* ARIMA-SVR.

3.5. Peramalan Harga Emas Menggunakan Model Hybrid ARIMA-SVR

Tahapan analisis selanjutnya adalah membentuk model *hybrid* ARIMA-SVR dengan menggabungkan komponen linear dan komponen nonlinear. Komponen linear merupakan hasil peramalan model ARIMA terbaik sedangkan komponen nonlinear merupakan hasil peramalan residual ARIMA terbaik menggunakan model SVR. Pada penelitian ini diperoleh model *hybrid* ARIMA-SVR merupakan kombinasi model ARIMA (2,1,2) dan model SVR menggunakan kernel *Radial Basic Function* (RBF) dengan nilai parameternya adalah $C = 2^{-0,5}$, $\gamma = 2^{8,25}$ dan $\varepsilon = 0,02$.

Langkah selanjutnya adalah melihat sejauh mana akurasi atau ketepatan model *hybrid* ARIMA-SVR untuk peramalan harga emas dengan membandingkan akurasi peramalan menggunakan model ARIMA. Penentuan keakuratan model dan peramalan dilakukan dengan membandingkan nilai MAPE. Hasil perhitungan nilai MAPE pada data *training* dan data *testing* pada model ARIMA dan *hybrid* ARIMA-SVR disajikan pada Tabel 7.

Tabel 7. Perbandingan MAPE model ARIMA dan *hybrid* ARIMA-SVR

Model	MAPE (%)	
	<i>Training</i>	<i>Testing</i>
ARIMA	0,903	4,076
<i>Hybrid</i> ARIMA-SVR	0,355	4,001

Tabel 7 menunjukkan bahwa nilai MAPE model *hybrid* ARIMA-SVR sebesar 0,355 pada data *training* dan 4,001 pada data *testing*. Lebih kecil dari nilai MAPE model ARIMA sebesar 0,903 pada data *training* dan 4,076 pada data *testing*. Sehingga disimpulkan bahwa model *hybrid* ARIMA-SVR secara signifikan mampu memberikan nilai peramalan lebih akurat dibandingkan model ARIMA.

Tahap selanjutnya adalah menggunakan model *hybrid* ARIMA-SVR untuk meramalkan harga emas untuk 12 hari kedepan yaitu pada 1 Desember 2020 sampai 14 Desember 2020. Hasil peramalan harga emas pada 1 Desember 2020 sampai 14 Desember 2020 dapat dilihat pada Tabel 8.

Tabel 8. Peramalan harga emas 1 Desember 2020 - 14 Desember 2020

Tanggal	Harga emas (Rp/gram)
1 Desember 2020	882.944
2 Desember 2020	882.693
3 Desember 2020	887.251
4 Desember 2020	882.241
5 Desember 2020	886.377
7 Desember 2020	882.892
8 Desember 2020	884.054
9 Desember 2020	883.238
10 Desember 2020	882.899
11 Desember 2020	884.439
12 Desember 2020	883.100
14 Desember 2020	882.891

4. Kesimpulan

Prosedur membangun model *hybrid* ARIMA-SVR adalah menggunakan model ARIMA untuk membentuk komponen linear. Residual dari model ARIMA dijadikan input untuk model SVR. Hasil pemodelan residual ARIMA menggunakan model SVR membentuk komponen nonlinear. Model *hybrid* ARIMA-SVR merupakan kombinasi komponen linear dan komponen linear tersebut. Model *hybrid* ARIMA-SVR terbaik untuk peramalan data harga emas adalah kombinasi model ARIMA (2,1,2) dan model SVR menggunakan kernel RBF dengan nilai parameternya adalah $C = 2^{-0,5}$, $\gamma = 2^{8,25}$ dan $\varepsilon = 0,02$. Model *hybrid* ARIMA-SVR ini dipilih karena memiliki nilai peramalan lebih akurat daripada model ARIMA, yang ditunjukkan dari nilai MAPE lebih kecil baik untuk data *training* maupun *testing*.

Referensi

- [1] N. U. Habibah, "Perkembangan Gadai Emas Ke Investasi Emas Pada Pegadaian Syariah," *Amwaluna : Jurnal Ekonomi dan Keuangan Syariah*, vol. 1, no. 1, 81-97, 2017.
- [2] G. E. Box, and G. M. Jenkins, *Time Series Analysis Forecasting and Control*, California: Holden Day, 1976.
- [3] Y. S. Kao, K. Nawata, and C. Y. Huang, "Predicting Primary Energy Consumption Using Hybrid ARIMA and GA-SVR Based on EEMD Decomposition", *Mathematics* 8, 1722, 2020.
- [4] B. E. Boser, I. Guyon, and V. Vapnik, "A Training Algorithm for Optimal Margin

- Classifiers", in *Proceedings of the fifth annual workshop on Computational learning theory (COLT)*, 1992, 144-152.
- [5] C. W. Hsu, C. C. Chang, and C. J. Lin, *A Practical Guide to Support Vector Classification*, Taipei: Department of Computer Science National Taiwan University, 2016.
- [6] J. D. Cryer, *Time Series Analysis*, Boston: PWS-KENT Publishing Company, 1986.
- [7] J. B. Ramsey, "Tests for Specification Errors in Classical Linear Least Squares Regression Analysis", *Journal of the Royal Statistical Society, Series B*, 31 (2), 350-371, 1969.
- [8] W.W.S. Wei, *Time Series Analysis: Univariate and Multivariate Methods*, Second Edition, New York: Pearson Education, 2006.
- [9] S. Haykin, *Neural Networks and Learning Machines*, Third Edition, New Jersey: Pearson Education, Inc, 2009.
- [10] A. J. Smola and B. Scholkopf, "A Tutorial on Support Vector Regression", *Statistic and Computing* 14, 199-222, 2004.
- [11] B. Santosa, *Data Mining: Teknik Pemanfaatan Data untuk Keperluan Bisnis, Teori dan Aplikasi*, Yogyakarta: Graha Ilmu, 2007.
- [12] V. N. Vapnik, *The Nature of Statistical Learning Theory*, New York: Springer, 1995.
- [13] S. Yadav, and S. Shukla, "Analysis of k-Fold Cross-Validation over Hold-Out Validation on Colossal Datasets for Quality Classification", *IEEE 6th International Conference on Advanced Computing (IACC)*, 2016.
- [14] Y. Zhang, L. Luo, J. Yang, D. Liu, R. Kong, and Y. Feng, "A hybrid ARIMA-SVR approach for forecasting emergency patient flow", *Journal of Ambient Intelligence and Humanized Computing* 10(3), 2018.
- [15] S. S. Balkin, and J. K. Ord, "Automatic Neural Network Modeling for Univariate Time Series", *International Journal of Forecasting*, Vol. 16: 509-515, 2000.



This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/). Editorial of JJoM: Department of Mathematics, Universitas Negeri Gorontalo, Jln. Prof. Dr. Ing. B.J. Habibie, Moutong, Tilongkabila, Kabupaten Bone Bolango, Provinsi Gorontalo 96119, Indonesia.