ISSN: 2339-2541
JURNAL GAUSSIAN, Volume 11, Nomor 4, Tahun 2022, Halaman 512 - 521
Online di: https://ejournal3.undip.ac.id/index.php/gaussian/

ANALISIS *SUPPORT VECTOR REGRESSION* (SVR) DENGAN ALGORITMA *GRID SEARCH TIME SERIES CROSS VALIDATION* UNTUK PREDIKSI JUMLAH KASUS TERKONFIRMASI COVID-19 DI INDONESIA

Anindita Nur Safira[1*], Budi Warsito[2], Agus Rusgiyono[3]
[1,2,3] Departemen Statistika, Fakultas Sains dan Matematika, Universitas Diponegoro
***e-mail :** aninditasafira10@gmail.com

DOI: 10.14710/j.gauss.11.4.512-521

Article Info:
Received: 2022-08-12
Accepted: 2022-10-20
Available Online: 2023-02-25

Keywords:
Support Vector Regression (SVR); *Grid Search* Algorithm; *Time Series Cross Validation*; Kernel; MAPE (*Mean Absolute Percentage Error*)

Abstract: *Coronavirus Disease* 2019 or Covid-19 is a group of types of viruses that interfere with the respiratory tract associated with the seafood market that emerged in Wuhan City, Hubei Province, China at the end of 2019. The first confirmed cases of Covid-19 in Indonesia on March 2, 2020, were 2 cases and until the end of 2021, it continues to grow every day. The purpose of this study was to predict the number of confirmed cases of Covid-19 in Indonesia using the *Support Vector Regression* (SVR) method with linear kernel functions, radial basis functions (RBF), and polynomials. *Support Vector Regression* (SVR) is the application of a *support vector machine* (SVM) in regression cases that aims to find the dividing line in the form of the best regression function. The advantage of the SVR model is can be used on time series data, data that are not normally distributed and data that is not linear. Parameter selection for each kernel used a *grid search* algorithm combined with *time series cross validation*. The criteria used to measure the goodness of the model are MSE (*Mean Square Error*), MAPE (*Mean Absolute Percentage Error*) and R^2 (Coefficient of Determination). The results of this study indicate that the best model is Support Vector Regression (SVR) with a polynomial kernel and the parameters used include Cost = 1, degree = 1, and coefficient = 0.1. The polynomial kernel SVR model produces a MAPE value of 0.4946215%, which means the model has very good predictive ability. The prediction accuracy obtained with an R2 value of 85.65011% and an MSE value of 161606.1.

1. PENDAHULUAN

Pada 31 Desember 2019, *World Health Organization* (WHO) mendapatkan laporan adanya 44 kasus pasien *pneumonia* yang tidak diketahui penyebabnya terdeteksi di Kota Wuhan, Provinsi Hubei, Cina. Pihak berwenang Cina mengidentifikasi adanya jenis virus *corona* baru yang menjadi penyebab kasus *pneumonia*. Wabah tersebut terus meluas hingga beberapa Negara di Dunia, sehingga pada 30 Januari 2020 WHO menetapkan status "*Global Emergency*" atau Pandemi Global pada 2019-nCoV. Pada 11 Februari 2020 WHO menamai penyakit tersebut dengan "*Coronavirus Disease* 2019" atau yang lebih dikenal dengan COVID-19 (WHO, 2020a).

Indonesia merupakan salah satu dari sekian banyak Negara yang juga melaporkan adanya kasus Covid-19. Kasus Covid-19 di Indonesia pertama kali dikonfirmasi pada 2 Maret 2020 sebanyak 2 kasus (WHO, 2020b). Kasus Covid-19 terus bertambah setiap harinya sehingga Indonesia menetapkan Covid-19 sebagai bencana nasional sejak tanggal 14 Maret 2020. Tercatat hingga 1 juli 2021 Indonesia menempati peringkat ke 5 sebagai Negara dengan jumlah kasus terkonfirmasi Covid-19 tertinggi di Dunia dengan total kasus 2.203.108 (worldometers, 2021). Upaya yang dilakukan pemetintah dalam pencegahan dan pengendalian Covid-19 diantaranya pemberlakuan Pembatasan Sosial Berskala Besar (PSBB), Pemberlakuan Pembatasan Kegiatan Masyarakat (PPKM), menerapkan protokol

kesehatan, *social distancing, new normal*, meniadakan kegiatan belajar mengajar di sekolah digantikan dengan belajar daring di rumah, menerapkan kerja dari rumah (*work from home*), dan juga melaksanakan vaksinasi Covid-19 (Nawangsari, 2021).

Metode yang digunakan untuk memprediksi jumlah kasus terkonfirmasi Covid-19 di Indonesia pada penelitian ini adalah metode *Support Vector Regression* (SVR). *Support Vector Regression* (SVR) merupakan penerapan *support vector machine* (SVM) yang digunakan pada kasus regresi. Metode SVR dapat digunakan pada data time series, data yang tidak berdistribusi normal dan data yang tidak linier. Penerapan metode SVR perlu dikombinasikan dengan suatu *loss function* dan juga fungsi kernel. Pada penelitian ini loss function yang digunakan adalah ε *-Insensitive loss function.* ε*-insensitive loss function* adalah fungsi yang menentukan tawar menawar (*trade off*) antara ketipisan fungsi (*flatness of function*) $f(\mathbf{x})$ dan batas toleransi terhadap residual. ε didefinisikan sebagai batas toleransi terhadap residual (Prahutama et al., 2014).

Fungsi kernel yang dapat digunakan pada metode SVR adalah fungsi kernel linier, *radial basis function* (RBF), dan polinomial (Prahutama et al., 2014). Tujuan yang akan dicapai oleh metode SVR menggunakan ε *-Insensitive loss function* adalah untuk menemukan fungsi $f(x)$ pemisah (*hyperplane*) yang mempunyai ε paling besar dari target aktual untuk keseluruhan data latih dan pada saat yang sama juga dicari fungsi yang setipis mungkin. Pemilihan parameter pada model SVR kernel linier, radial basis function, dan polinomial berupa ε (Epsilon), C (Cost), σ (Sigma), degree, dan koefisien menggunakan algoritma *Grid Search Time Series Cross Validation*. Penentuan prediksi terbaik dilakukan dengan melihat kriteria keakuratan prediksi yang pada penelitian ini meliputi MSE (*Mean Square Error*), MAPE (*Mean Absolute Percentage Error*) dan R^2 (Koefisien Determinasi).

2. TINJAUAN PUSTAKA

Coronavirus Disease 2019 atau Covid-19 merupakan kelompok jenis virus yang menganggu saluran pernapasan dan salah satunya pernah menyebabkan munculnya wabah *Severe Acute Respiratory Infection* (SARS) di dunia (Abdillah et al., 2020). Covid-19 ditetapkan sebagai pandemi global atau "*Global Emergency*" oleh *World Health Organization* (WHO) pada tanggal 30 Januari 2020 (WHO, 2020a). Dengan ditetapkannya status pandemi, WHO mengimbau seluruh negara terdampak untuk membuat maupun menjadi fokus untuk mengenali, menguji, menjaga, mengisolasi, melacak, serta memobilisasi warga (Abdillah et al., 2020).

Time Series merupakan sekumpulan data yang diobservasi selama kurun waktu tertentu yang terurut secara kronologis. Dasar pemikiran *Time Series* adalah pengamatan sekarang(Z_t) bergantung pada satu atau beberapa pengamatan sebelumnya(Z_{t-n}). Model *Time Series* dibuat karena secara statistik terdapat korelasi antar deret pengamatan yang sering dikenal dengan *Autocorellation Function* (ACF). Komponen input (x) didapatkan dengan mencari nilai lag dari plot PACF yang terbentuk pada proses autoregresif yang artinya data berregresi dengan dirinya sendiri (Rianto & Yunis, 2021).

Regresi linier (*linear regression*) merupakan suatu teknik statistika yang menghasilkan suatu persamaan linier. Tujuan analisis regresi adalah untuk menentukan model statistik (dalam bentuk formula matematik) yang dapat digunakan untuk melakukan prediksi nilai-nilai variabel terikat (disebut juga variabel respon) Y berdasarkan nilai-nilai dari variabel-variabel bebas (disebut juga variabel prediktor) $X_1,X_2,....,X_k$.

Support Vector Regression (SVR) merupakan penerapan SVM yang digunakan pada kasus regresi. Tujuan yang akan dicapai pada SVR adalah untuk menemukan fungsi $f(x)$

pemisah (*hyperplane*) yang mempunyai ε yang paling besar dari target aktual y_i untuk keseluruhan data latih dan pada saat yang sama juga dicari fungsi yang setipis mungkin. Semua kesalahan (selisih antara prediksi dengan aktual) akan diabaikan jika bernilai kurang dari ε dan jika lebih besar dari ε maka akan dikalikan dengan C (Saputra et al., 2019).

Pada metode SVR fungsi regresi $f(\mathbf{x})$ dapat dinyatakan dengan formula sebagai berikut (Scholkopf & Smola, 2004) :

$$f\left(\mathbf{x}\right) = \mathbf{w}^T \varphi\left(\mathbf{x}\right) + \mathbf{b} \tag{1}$$

dengan $\mathbf{w}$ adalah vektor pembobot berdimensi ℓ ; $\varphi\left(\mathbf{x}\right)$ = fungsi yang memetakan x pada ruang dengan ℓ dimensi; dan $\mathbf{b}$ = vektor bias konstan. Penyelesaian problem optimasi *hyperlane* dilakukan dengan bentuk *Quadratic Programming* sebagai berikut:

$$\min \frac{1}{2}\|\mathbf{w}\|^2 \tag{2}$$

dengan syarat

$$\begin{aligned} &\mathbf{y}_i - \mathbf{w}^T \varphi(\mathbf{x}_i) - \mathbf{b} \le \boldsymbol{\varepsilon} \\ &\mathbf{w}^T \varphi(\mathbf{x}_i) - \mathbf{y}_i + \mathbf{b} \le \boldsymbol{\varepsilon} \\ &i = 1, 2, ..., \ell \end{aligned} \tag{3}$$

dengan $\mathbf{y}_i$ merupakan nilai aktual data ke I dan $\boldsymbol{\varepsilon}$ merupakan vektor ε konstan. Faktor $\|\mathbf{w}\|^2$ disebut dengan *regulasi* dengan rumus $\|\mathbf{w}\|^2 = \left\|\mathbf{w}^* - \mathbf{1}\right\|^2$ dan $\mathbf{w}^* = \left(\mathbf{1} - \mathbf{w}^T\right)^T$. Meminimalkan $\|\mathbf{w}\|^2$ akan membuat suatu fungsi setipis (*flat*) mungkin, sehingga mampu mengontrol kapasitas fungsi (*function capacity*). Sebuah fungsi $f(\mathbf{x})$ diasumsikan dapat membaca semua titik $(\mathbf{x}_i, \mathbf{y}_i)$ dengan batas toleransi residual sebesar ε. Semua titik berada pada $f(\mathbf{x}) \pm \boldsymbol{\varepsilon}$ (*feasible*) dan apabila terdapat beberapa titik yang mungkin keluar dari $f(\mathbf{x}) \pm \boldsymbol{\varepsilon}$ (*infeasible*) maka perlu ditambahkan variabel slack ξ, ξ^* untuk mengatasi permasalahan pembatas yang tidak layak (*infeasible constraint*). Selanjutnya problem optimasi pada persamaan (2) dapat diformulasikan sebagai berikut :

$$\min \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{l}(\xi_i, \xi_i^*) \tag{4}$$

dengan syarat

$$\begin{aligned} &\mathbf{y}_i - \mathbf{w}^T \varphi(\mathbf{x}_i) - \mathbf{b} - \xi_i \le \varepsilon, i = 1, 2, ..., \ell \\ &\mathbf{w}^T \varphi(\mathbf{x}_i) - \mathbf{y}_i + \mathbf{b} - \xi_i^* \le \varepsilon, i = 1, 2, ..., \ell \\ &\xi_i, \xi_i^* \ge 0 \end{aligned} \tag{5}$$

Loss Function adalah sebuah fungsi yang digunakan untuk menghitung perbedaan antara nilai prediksi dan nilai aktual. Perbedaan dalam penggunaan *loss function* akan menghasilkan formula SVR yang berbeda (Santosa, 2007). *Loss function* yang dapat digunakan untuk data non linier adalah *ε-insensitive loss function* sebagai sebuah pendekatan

Huber's loss function yang memungkinkan serangkaian *support vector* akan diperoleh. Formulasi ε-*insensitive loss function* adalah sebagai berikut (Gunn, 1998):

$$L_{\varepsilon} = \begin{cases} 0, \text{untuk} \left| f(\mathbf{x}) - \mathbf{y} \right| < \varepsilon \\ \left| f(\mathbf{x}) - \mathbf{y} \right| - \varepsilon, \text{lainnya} \end{cases} \tag{6}$$

dengan L_{ε} adalah ε-*insensitive loss function*. Solusi optimasi *hyperlane* dapat diselesaikan dengan fungsi *Lagrange* dengan membangun dua set variabel sebagai berikut (Scholkopf & Smola, 2004):

$$L(\mathbf{w}, \mathbf{b}, \xi, \xi^*, \alpha, \alpha^*, \eta, \eta^*) = \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{\ell}\left(\xi_i + \xi_i^*\right) - \sum_{i=1}^{\ell}\alpha_i\left(\varepsilon + \xi_i - \mathbf{y}_i + w^T\varphi\left(\mathbf{x}_i\right) + \mathbf{b}\right)$$
$$-\sum_{i=1}^{\ell}\alpha_i^*(\varepsilon + \xi_i^* - \mathbf{y}_i + \mathbf{w}^T\varphi(\mathbf{x}_i) - \mathbf{b}) - \sum_{i=1}^{\ell}(\eta_i\xi_i + \eta_i^*\xi_i^*) \tag{7}$$

dengan $\alpha_i, \alpha_i^*, \eta_i, \eta_i^*$ adalah *lagrange multiplier*. Untuk mencari parameter optimasi *hyperlane*, fungsi L diturunkan parsial terhadap $\mathbf{w}, b, \xi, \xi*$. Hasil turunan dari fungsi L sebagai berikut:

$$\mathbf{w} = \sum_{i=1}^{\ell}\left(\alpha_i - \alpha_i^*\right)\varphi\left(\mathbf{x}_i\right) \tag{8}$$

$$C = \alpha_i + \eta_i \tag{9}$$

$$C = \alpha_i^* + \eta_i^* \tag{10}$$

Berdasarkan persamaan (24) fungsi $f(\mathbf{x})$ dapat ditulis sebagai:

$$f\left(\mathbf{x}\right) = \mathbf{w}^T\varphi(\mathbf{x}) + \mathbf{b}$$
$$f\left(\mathbf{x}\right) = \sum_{i=1}^{\ell}(\alpha_i - \alpha_i^*)\varphi^T(\mathbf{x}_i)\varphi\left(\mathbf{x}\right) + \mathbf{b} \tag{11}$$

Untuk mendapatkan solusi *b* digunakan *Karush-Kuhn-Tucker* (KKT) sehingga diperoleh estimasi (bias) sebagai berikut:

$$\mathbf{b} = \mathbf{y}_i - \mathbf{w}^T\varphi(\mathbf{x}_i) + \boldsymbol{\varepsilon} \text{ untuk } 0 < a_i < C \tag{12}$$

$$\mathbf{b} = \mathbf{y}_i - \mathbf{w}^T\varphi(\mathbf{x}_i) - \boldsymbol{\varepsilon} \text{ untuk } 0 < a_i^* < C \tag{13}$$

Menggunakan fungsi kernel, metode *Support Vector Regression* SVR dapat digunakan untuk menyelesaikan kasus nonlinier dengan cara dipisahkan secara linier dan dipetakan pada *feature space* yang baru. Terdapat 3 jenis kernel yang digunakan pada metode SVR sebagai berikut (Prahutama et al., 2014) :

1) Kernel Linier

$$K\left(\mathbf{x}_i, \mathbf{x}\right) = \mathbf{x}_i\mathbf{x}^T \tag{14}$$

2) Kernel Polinomial

$$K\left(\mathbf{x}_i, \mathbf{x}\right) = \left(\mathbf{x}_i\mathbf{x}^T + r\right)^d \tag{15}$$

3. Kernel Radial Basis Function (RBF)

$$K\left(\mathbf{x}_i, \mathbf{x}\right) = \exp\left\{-\frac{1}{2\sigma^2}\left(\left\|\mathbf{x} - \mathbf{x}_i\right\|^2\right)\right\} \tag{16}$$

Cross-validation adalah sebuah teknik validasi model untuk menilai bagaimana hasil statistik analisis akan menggeneralisasi kumpulan data independent dimana data dipisahkan menjadi dua subset yaitu data proses (data latih) dan data evaluasi (data uji). Teknik ini

utamanya digunakan untuk melakukan prediksi model dan memperkirakan seberapa akurat sebuah model prediktif ketika dijalankan dalam praktiknya. *Cross validation* yang dapat digunakan pada *time series* adalah *Time Series Cross Validation*.

Pada *time series cross validation*, *training set* (data latih) hanya terdiri dari observasi yang terjadi sebelum observasi yang membentuk *test set* (data uji). Hal tersebut mengakibatkan tidak terdapat pengamatan masa depan yang dapat digunakan dalam menyusun peramalan (Athanasopoulos, 2021).

Salah satu algoritma yang dapat digunakan untuk menentukan parameter optimal pada model SVR adalah algoritma *grid search*. Pada algoritma ini akan dilakukan pembagian jangkauan parameter yang akan dioptimalkan ke dalam *grid* dan melintasi semua titik untuk mendapatkan parameter optimal (Prahutama et al., 2014). Metode *grid search* dilakukan dengan mencoba kombinasi parameter satu persatu dan membandingkan nilai terbaik yang diberikan oleh parameter tersebut. Dalam aplikasinya, algoritma *grid search* harus dipandukan dengan beberapa metrik kinerja, biasanya diukur dengan *cross-validation* pada data latih.

Dalam semua situasi prediksi pasti mengandung derajat ketidakpastian yang sebut dengan *error* (unsur kesalahan). Sumber penyimpangan dalam prediksi bukan hanya disebabkan oleh unsur *error*, tetapi ketidakmampuan suatu model prediksi mengenali unsur yang lain dalam deret data juga mempengaruhi besarnya penyimpangan dalam prediksi. Untuk mengevaluasi akurasi dan prediksi kinerja model berbeda, penelitian ini mengadopsi tiga indeks evaluasi:

1) *Mean Square Error* (MSE)

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{17}$$

2) *Mean Absolute Percentage Error* (MAPE)

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}|PE_i| = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{(y_i-\hat{y}_i)}{y_i}\right| x100 \tag{18}$$

3) Koefisien determinasi (R^2)

$$R^2 = 1 - \frac{JKE}{JKT} \tag{19}$$

dengan JKE dan JKT sebagai

$$\begin{aligned} \text{JKE} &= \sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2 \\ \text{JKT} &= \sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2 \end{aligned} \tag{20}$$

3. METODE PENELITIAN

Jenis data yang digunakan dalam penelitian ini adalah data sekunder yaitu data jumlah kasus terkonfirmasi Covid-19 di Indonesia yang diambil dari situs resmi (https://covid19.go.id) dari tanggal 1 Januari 2021 sampai dengan 30 November 2021 dengan jumlah data sebanyak 334 data.

Data jumlah kasus terkonfirmasi Covid-19 di Indonesia tanggal 1 Januari hingga 30 November 2021 dimodifikasi menjadi variabel prediktor (x) dan variabel respon (y) melalui plot PACF. Langkah awal untuk menentukan variabel prediktor (x) adalah membuat plot PACF pada data kasus terkonfirmasi Covid-19 di Indonesia. Dengan mencari nilai lag dari plot PACF yang terbentuk pada proses autoregresif diperoleh variabel predictor (x). Data jumlah kasus terkonfirmasi Covid-19 di Indonesia berupa data historis kemudian

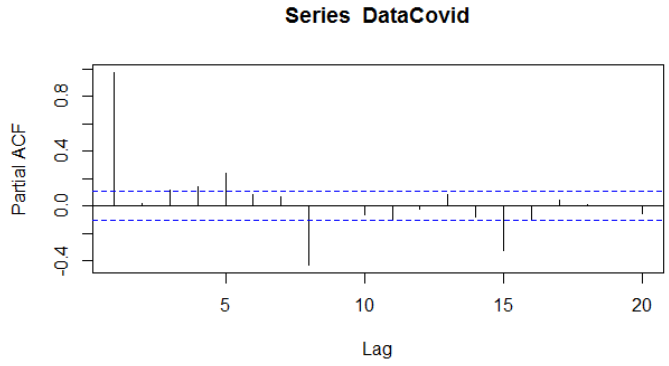
berlandaskan rumus *time series* data dimodifikasi menjadi data ke-1, …, n-1 sebagai variabel prediktor (x) dan data ke-2, …, n sebagai variabel respon (y).

Pengolahan data dalam penelitian ini menggunakan bantuan *software Rstudio*. Langkah-langkah analisis datanya adalah sebagai berikut :

1. Mengumpulkan data Jumlah kasus terkonfirmasi Covid-19 di Indonesia periode Januari hingga November 2021
2. Memodifikasi data Jumlah kasus terkonfirmasi Covid-19 di Indonesia periode Januarit hingga November 2021 menjadi variabel prediktor (x) dan variabel respon (y) melalui plot PACF.
3. Membagi data menjadi data latih dan data uji dengan proporsi tertentu.
4. Melakukan analisis SVR untuk menghasilkan pemodelan menggunakan *softwere RStudio* yang terdiri :
   a. Memilih parameter terbaik menggunakan algoritma *grid search time series cross validation*.
   b. Melakukan analisis *support vector regression* (SVR) dengan parameter terbaik, fungsi *ε-insensitive Loss Function* dan Fungsi Kernel Linier, Polinomial, dan Radial.
   c. Melakukan prediksi data dengan menggunakan model *support vector regression* (SVR) dengan parameter terbaik, fungsi *ε-insensitive Loss Function* dan Fungsi Kernel Linier, Polinomial, dan Radial.
5. Menguji Model dengan menggunakan MSE (*Mean Square Error*), MAPE (*Mean Absolute Percentage Error*), R^2 (Koefisien Determinasi).

4. HASIL DAN PEMBAHASAN

Mencari nilai lag dari plot PACF yang terbentuk pada proses autoregresif diperoleh komponen input, yang artinya data berregresi dengan dirinya sendiri (Rianto & Yunis, 2021). Identifikasi lag sebagai komponen input didasarkan pada koefisien autokorelasi parsial terputus atau signifikan.



Gambar 1 Plot PACF

Gambar 1 menunjukkan bahwa diperoleh 6 variabel lag yang akan dijadikan sebagai komponen input, yaitu lag 1, 3, 4, 5, 8, dan 15. Dengan demikian pada penelitian ini variabel respon (y) adalah jumlah kasus terkonfirmasi Covid-19 di Indonesia, dan variabel prediktor (x) adalah Z_{t-1}, Z_{t-3}, Z_{t-4}, Z_{t-5}, Z_{t-8}, dan Z_{t-15}.

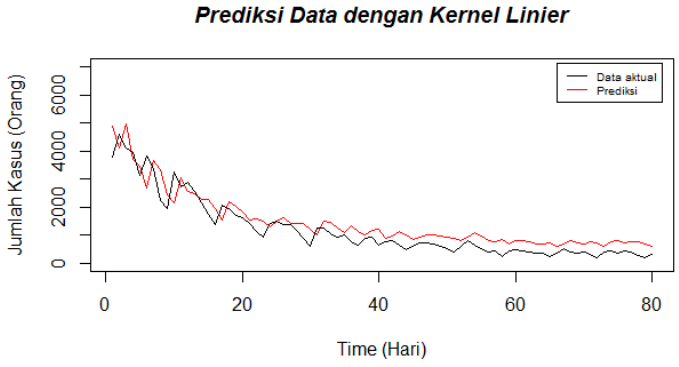
Data sebanyak 334 dibagi menjadi dua dengan data latih sebanyak 80% atau 254 data dan data latih sebanyak 20% atau 80 data. Pada proses pembentukan model dan penentuan parameter terbaik dengan *grid search time series cross validation* digunakan data latih.

Pada proses validasi menggunakan *time series cross validation* data latih akan dibagi menjadi data latih dan data validasi untuk melihat seberapa akurat model yang terbentuk. Data latih yang digunakan pada proses ini dimulai pada 14 data pertama hingga data 174. Pemilihan data latih sebanyak 14 data pertama karena 254 data dikurangkan dengan kelipatan 80, agar jumlah data validasi tetep yaitu 80 data. Pada iterasi pertama data latih yang digunakan adalah 14 data pertama dan data validasi yang digunakan adalah 80 data setelah data latih. Pada iterasi kedua data latih yang digunakan adalah 15 data pertama dan data validasi adalah 80 data setelah data latih, begitu seterusnya hingga iterasi terakhir. Pada penelitian ini iterasi yang dilakukan sebanyak 161 iterasi.

Tahap awal yang perlu dilakukan untuk melakukan prediksi dengan menggunakan metode SVR adalah menentukan nilai parameter. Parameter yang digunakan pada kernel linier adalah ε (Epsilon) dan C (Cost), yang dihasilkan dengan menggunakan algoritma *grid search* yang dipadukan dengan metode *time series cross validation* pada data latih dan data validasi dengan program *RStudio*. Parameter yang digunakan merupakan parameter terbaik untuk *hyperplane*, yang ditentukan dengan melihat nilai Root Mean Square Error (RMSE) terkecil.

Nilai ε pada penelitian ini bersifat konstan yaitu 0.1 yang diambil dari *default* pada program *RStudio*. Nilai percobaan C adalah 0.25, 0.5, dan 1 (Prahutama et al., 2014). Perhitungan *tuning* parameter menggunakan metode *grid search time series cross validation* pada kernel linier diperoleh bahwa nilai C adalah konstan yaitu C = 1. Berdasarkan Tabel 7 diperoleh bahwa nilai Cost atau C adalah konstan yaitu C = 1. Tahap selanjutnya adalah menghitung nilai dari β (beta) dan b (bias). Nilai β dihitung melalui fungsi *langrange multiplier* dan nilai bias didapatkan dengan menggunakan *Karush-Kuhn Tucher* pada penelitian ini dilakukan dengan bantuan program *RStudio*. Nilai bias pada model SVR dengan kernel linier sebesar b = 0.00919163. Banyaknya nilai β bergantung pada banyaknya *support vector*. Pada model Pada model SVR dengan kernel linier didapatkan *support vector* sebanyak 95 data.

Setelah didapatkan parameter terbaik, banyaknya *support vector*, nilai β dan b maka dapat dilakukan prediksi pada data uji.



Gambar 2 Plot data prediksi model SVR kernel linier

Gambar 2 memperlihatkan bahwa data aktual dibandingkan dengan data hasil prediksi memiliki pola yang sama, sehingga dapat dikatakan bahwa model SVR dengan kernel linier dapat mengatasi masalah *overfitting* yang dalam memprediksi jumlah kasus terkonfirmasi Covid-19 di Indonesia.

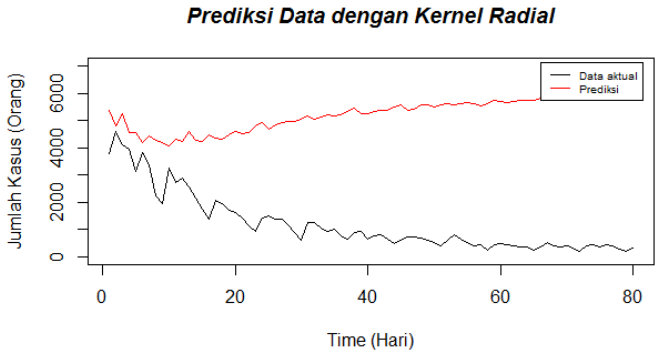
Sama halnya dengan pembentukan SVR kernel linier, tahap awal yang perlu Sama halnya dengan pembentukan SVR kernel linier, tahap awal yang perlu dilakukan adalah

menentukan nilai parameter. Parameter yang diperlukan pada SVR kernel *radial basis function* (RBF) adalah parameter ε (epsilon), C (Cost) dan σ (Sigma). Nilai ε pada penelitian ini bersifat konstan yaitu 0.1 yang diambil dari *default* pada program *RStudio*. Nilai percobaan C adalah 0.25, 0.5, dan 1. Nilai σ percobaan yaitu 4 dan 5 (Saputra et al., 2019). Perhitungan *tuning* parameter menggunakan metode *grid search time series cross validation* pada kernel radial output sebagai berikut:

Tabel 1 *Grid Search time series cross validation* kernel radial

C	Sigma	RMSE	R^2	MAE
0.25	5.872546	11142.71	0.07966043	8797.200
0.50	5.872546	11007.74	0.08200792	8664.417
1.00	5.872546	10868.06	0.08504678	8539.163

Berdasarkan Tabel 1 nilai C dan sigma terbaik pada kernel *radial basis function* yaitu C = 1 dengan sigma konstan yaitu 5.872546. Nilai b dapat dicari menggunakan bantuan paket program *RStudio*, sehingga didapatkan bahwa nilai bias sebesar b = - 1.488849. Banyaknya *support vector* pada SVR kernel *radial basis function* (RBF) didapatkan sebanyak 106 data.



Gambar 3 Plot data prediksi model SVR kernel radial

Dapat dilihat pada Gambar 3 bahwa grafik data aktual dibandingkan dengan hasil prediksi memiliki pola yang prediksi tidak memiliki pola yang sama, sehingga dapat dikatakan bahwa model SVR dengan kernel radial kurang dapat mengatasi masalah *overfitting* yang dalam memprediksi jumlah kasus terkonfirmasi Covid-19 di Indonesia. Pada analisis yang telah dilakukan pada model SVR kernel radial didapatkan jumlah *support vector* yang lebih banyak dibandingkan dengan kernel lain. Jumlah *support vector* sebanyak 106 data artinya terdapat 106 data yang terletak pada dan diluar batas dari fungsi keputusan. Hal tersebut dapat diminimalisir dengan cara menaikkan nilai epsilon agar jumlah *support vector* menurun.

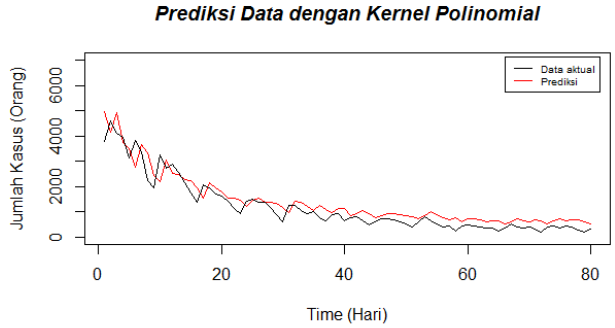
Seperti yang dilakukan pada pembentukan SVR kernel linier dan kernel radial yang perlu dilakukan adalah menentukan nilai parameter ε (Epsilon), C (Cost), degree, dan koefisien pada pembentukan SVR kernel polinomial. Nilai ε pada penelitian ini bersifat konstan yaitu 0.1 yang diambil dari *default* pada program *RStudio*. Nilai percobaan C adalah 0.25, 0.5, dan 1. Nilai percobaan degree yaitu 1,2, dan 3. Nilai percobaan koefisien meliputi 0.001, 0.010, dan 0.100 (Adiningtyas et al., 2015).

Penentuan nilai parameter dilakukan dengan menggunakan algoritma *grid search time series cross validation*. Berikut merupakan hasil perhitungan tuning parameter pada kernel polinomial:

Tabel 2 *Grid Search time series cross validation* kernel polinomial

Degree	Scale	Cost	RMSE	R^2	MAE
1	0.001	0.25	10192.281	0.6533934	8069.093
1	0.001	0.50	9161.992	0.6551626	7220.480
1	0.001	1.00	7653.539	0.6580895	5962.803
1	0.010	0.25	5667.574	0.6661357	4370.596
1	0.010	0.50	4538.005	0.6763360	3532.464
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
3	0.100	1.00	158504.192	0.4361690	85464.733

Berdasarkan Tabel 2 diperoleh nilai C, degree, dan koefisien terbaik yaitu C = 1, degree=1, dan koefisien=0.1. Nilai parameter b atau bias pada model SVR dengan kernel polinomial sebesar b = 0.006085192. SVR dengan kernel polinomial didapatkan *support vector* sebanyak 96 data.



Gambar 4 Plot data prediksi model SVR kernel polinomial

Berdasarkan Gambar 4 diketahui bahwa grafik data actual dibandingkan dengan hasil prediksi memiliki pola yang sama, sehingga dapat dikatakan bahwa model SVR dengan kernel polinomial dapat mengatasi masalah *overfitting* yang dalam memprediksi jumlah kasus terkonfirmasi Covid-19 di Indonesia.

Dalam penelitian ini perhitungan akurasi atau ketepatan prediksi menggunakan R^2 sedangkan perhitungan *error* atau kesalahan prediksi menggunakan MSE dan MAPE.

Tabel 4 Perbandingan nilai MSE, MAPE, R^2

Model	MSE	MAPE	R^2
SVR kernel linier	198693.1	0.6095872%	82.35695%
SVR kernel radial basis function	18478525	7.81219%	54.27522%
SVR kernel polinomial	161606.1	0.4946215%	85.65011%

Berdasarkan Tabel 4 terlihat bahwa Model SVR kernel polinomial lebih akurat dalam melakukan prediksi jumlah kasus terkonfirmasi Covid-19 di Indonesia dibandingkan dengan model SVR kernel radial basis function maupun SVR kernel polinomial.

5. KESIMPULAN

Berdasarkan hasil analisis pembahasan yang telah dilakukan dapat diambil kesimpulan bahwa model SVR kernel polinomial adalah model terbaik dalam memprediksi jumlah kasus

terkonfirmasi Covid-19 di Indonesia. Parameter yang digunakan pada model SVR ini adalah C = 1, degree=1, dan koefisien=0.1. Hasil prediksi yang dihasilkan oleh model SVR kernel polinomial memiliki pola yang sama dengan data aktual, dengan kata lain model SVR dapat mengatasi masalah *overfitting* dalam memprediksi jumlah kasus terkonfirmasi Covid-19 di Indonesia. Prediksi jumlah terkonfirmasi Covid-19 di Indonesia menggunakan metode *Support Vector Regression* (SVR) dengan kernel polinomial menghasilkan nilai MSE sebesar 161606.1 dan nilai MAPE sebesar 0.4946215% yang artinya akurasi prediksi sangat baik atau model mempunyai kemampuan prediksi yang sangat baik. Nilai R^2 sebesar 0.8565011 atau 85.65011% yang artinya prediksi data jumlah terkonfirmasi Covid-19 di Indonesia menggunakan metode SVR dengan fungsi kernel polinomial mempunyai hasil akurasi atau ketepatan prediksi yang sangat baik.

DAFTAR PUSTAKA

Abdillah, L. A., Faried, A. I., Febrianty, Iqbal, M., Masrul, Mastuti, R., Napitupulu, D., Prianto, C., Puji Hastuti, J., Purba, D. W., Purnomo, A., Rahmadana, M. F., Ramadhani, Y. R., Saputra, D. H., Sari, J. D. C., Simarmata, J., Sulaiman, D. O. K., Soetijono, I. K., Tasnim, & Vinolina, N. S. (2020). *Pandemik COVID-19: Persoalan dan Refleksi di Indonesia* (T. Limbong (ed.)). Yayasan Kita Menulis.

Adiningtyas, D. T., Mukid, M. A., & Safitri, D. (2015). Peramalan Jumlah Tamu Hotel Di Kabupaten Demak Menggunakan Metode Support Vector Regression. *None*, *4*(4), 785–794.

Athanasopoulos, R. J. H. and G. (2021). *Forecasting: Principles and Practice (3rd ed)*. Monash University, Australia. https://otexts.com/fpp3

Firmansyah, N. Y., Nawangsari, E. R., Rahmadani, A. W., & Zachary, Y. A. (2021). Partisipasi Masyarakat Kelurahan Jelakombo Terhadap Pemberlakuan Pembatasan Kegiatan Masyarakat (PPKM) Skala Mikro Di Kabupaten Jombang. *Jurnal Syntax Transformation*, *2*(5), 593–605.

Gunn, S. R. (1998). Support Vector Machines for classification and regression. In *Analyst*.

Prahutama, A., Utami, T. W., & Yasin, H. (2014). Prediksi Harga Saham Menggunakan Support Vector Regression Dengan Algoritma Grid Search. *Media Statistika*, *7*(1), 29–35.

Rianto, M., & Yunis, R. (2021). Analisis Runtun Waktu Untuk Memprediksi Jumlah Mahasiswa Baru Dengan Model Random Forest. *Paradigma - Jurnal Komputer Dan Informatika*, *23*(1).

Santosa, B. (2007). *Data Mining Terapan dengan MATLAB* (1st ed.). Graha Ilmu.

Saputra, G. H., Sartono, B., & Wigena, A. H. (2019). *Penggunaan Support Vector Regession dalam Pemodelan Indeks Saham Syariah Indonesia dengan Algoritma Grid Search*. 148–160.

Scholkopf, B., & Smola, A. J. (2004). A tutorial on support vector regression. *Statistics and Computing*, *14*, 199–222.

WHO. (2020a). 15-Novel Coronavirus (2019-nCoV). *World Health Organization*, *February*, 1–7.

WHO. (2020b). Coronavirus disease 2019 (COVID-19) Situation Report – 42 Data as reported by 10 AM CET 02 March 2020 H. *World Health Organization*, *14*(6), e01218.

worldometers. (2021). *Reported Cases and Deaths by Country or Territory*. Www.Worldometers.Info.https://www.worldometers.info/coronavirus/?zarsrc=130#main_table