

PERAMALAN KASUS HARIAN MONKEYPOX DUNIA DENGAN PENDEKATAN *SUPPORT VECTOR REGRESSION*

Marcel Laverda Subiyanto¹, Yulia Amanda², Muhammad Nadhil Fachrian³, Afriani⁴, Achmad Yazid Busthomi Rohim⁵, Nur Chamidah^{6*}

^{1,2,3,4,5,6}Departemen Matematika, Fakultas Sains dan Teknologi, Universitas Airlangga
e-mail: ¹marcel.laverda.subiyanto-2019@fst.unair.ac.id, ²yulia.amanda-2019@fst.unair.ac.id,
³muhammad.nadhil.fachrian-2019@fst.unair.ac.id, ⁴afriani-2019@fst.unair.ac.id,
⁵achmad.yazid.bustomi-2019@fst.unair.ac.id, ⁶nur-c@fst.unair.ac.id

*Penulis Korespondensi

ABSTRAK

Monkeypox saat ini menjadi perhatian masyarakat global. Maka, penting untuk mengetahui perkembangan jumlah kasus *monkeypox* kedepannya. Pada penelitian ini dilakukan peramalan kasus harian *monkeypox* menggunakan metode *Support Vector Regression* (SVR) dengan Fungsi *Kernel Radial Basis Function* (RBF). Data yang digunakan adalah data sekunder berupa deret waktu harian mulai 29 Mei sampai 20 Oktober 2022. Untuk memperoleh parameter optimal pada model SVR, peneliti menggunakan algoritma *grid search* untuk memprediksi data *testing* secara akurat. Berdasarkan hasil pencarian parameter optimal menggunakan algoritma *grid search* diperoleh parameter optimalnya adalah $C = 2^{0,6}$, $\gamma = 2^{0,2}$, dan $\varepsilon = 0,09$. Nilai RMSE untuk hasil peramalan menggunakan model SVR dengan parameter optimal pada data *training* dan *testing* sebesar 352,3 dan 809,7.

Kata kunci : Monkeypox, Peramalan, SVR, RBF, *Grid Search*

ABSTRACT

Monkeypox is a disease that becomes public health emergency of international concern. It is important to know the development of the number of monkeypox cases in the future. In this study, forecasting monkeypox daily cases was carried out using the Support Vector Regression (SVR) method with the Kernel Radial Basis Function (RBF). The data used is secondary data from May 29th to October 20th 2022. To obtain optimal parameters in the SVR model, researcher uses a grid search algorithm to accurately predict the testing data. Based on the search results for optimal parameters using the grid search algorithm, the optimal parameters are $C = 2^{0,6}$, $\gamma = 2^{0,2}$, and $\varepsilon = 0.09$. RMSE values for forecasting using the SVR model with optimal parameters on training and testing data are 352.3 and 809.7.

Keywords: Monkeypox, Forecasting, SVR, RBF, *Grid Search*

PENDAHULUAN

Salah satu penyakit endemi yang kini menjadi perhatian masyarakat global adalah penyakit *monkeypox* atau cacar monyet. Awal mula Virus *monkeypox* menginfeksi manusia adalah pada tahun 1970 di Republik Demokratik Kongo dimana virus tersebut dianggap sebagai penyakit endemik (Cabanillas *et al.*, 2022). Belum ditemukan alasan spesifik terjadinya penyebaran penularan penyakit *monkeypox* di daerah endemik dan non-endemik, akan tetapi hingga saat ini sebagian besar penularan terjadi melalui manusia ke manusia setelah melakukan perjalanan ke negara-negara Afrika. Penularan antar manusia terjadi melalui tetesan pernapasan, kontak dengan cairan tubuh, lingkungan atau barang-barang pasien yang terkontaminasi, lesi kulit orang yang terinfeksi dan dengan penularan seksual dari individu yang terinfeksi dengan lesi selangkangan dan genital (Jamil *et al.*, 2022).

Sejak awal Mei 2022, lebih dari 3000 infeksi virus cacar monyet telah dilaporkan di lebih dari 50 negara dari 5 benua mendorong *World Health Organization* (WHO) menyatakan cacar monyet sebagai keadaan darurat kesehatan global (Thornhill *et al.*, 2022). Pada tanggal 20 Juli 2022, dilaporkan bahwa kasus cacar monyet mencapai 14 ribu kasus di dunia oleh *World Health Organization* (WHO). Pada awal agustus 2022, kasus *monkeypox* kembali meningkat dimana lebih dari 26.500 kasus yang tersebar di 80 negara. Hingga Kamis, 18 Agustus 2022, dilaporkan bahwa kasus cacar monyet telah menyebar lebih dari 90 negara pada negara non-endemik dimana terdapat 38.800 kasus jumlah pasien yang telah terkonfirmasi *monkeypox* di dunia. (Farid, 2022).

Kasus pertama cacar monyet pertama kali muncul di Indonesia dilaporkan pada tanggal 20 Agustus 2022 (Wijaya, 2022). Pasien merupakan seorang laki-laki berusia 27 tahun asal Jakarta. Pasien tersebut pernah tercatat melakukan perjalanan ke beberapa negara pada 8 Agustus 2022. Pasien menunjukkan gejala demam, limpa membesar, dan ruam di wajah, telapak

tangan, kaki, dan bagian sekitar alat kelamin. Merespons munculnya kasus pertama cacar monyet di Indonesia, Ikatan Dokter Indonesia (IDI) telah membentuk gugus tugas yang didedikasikan untuk memerangi virus. Gugus tugas baru yang dibentuk mencakup hingga 25 dokter dari setidaknya enam organisasi medis yang berbeda, termasuk Perhimpunan Dokter Spesialis Penyakit Dalam Indonesia (PAPDI) dan Perhimpunan Dokter Spesialis Kulit dan Kelamin Indonesia (Perdoski) (Suhenda, 2022).

Berdasarkan penelitian yang dilakukan oleh John P. Thornhill *et al.*, (2022), melaporkan bahwa terdapat 528 infeksi yang didiagnosis antara 27 April dan 24 Juni 2022, di 43 lokasi di 16 negara. Secara keseluruhan, 98% orang yang terinfeksi adalah pria *gay* atau biseksual, 75% berkulit putih, dan 41% memiliki infeksi virus *human immunodeficiency*; dengan usia rata-rata adalah 38 tahun. Asumsi peneliti menyebutkan bahwa penularan kasus *monkeypox* terjadi melalui aktivitas seksual pada 95% orang yang terinfeksi. Infeksi menular seksual bersamaan dilaporkan pada 109 dari 377 orang (29%) yang diuji. Dalam rangkaian kasus ini, *monkeypox* dimanifestasikan dengan berbagai temuan klinis dermatologis dan sistemik. Identifikasi simultan kasus di luar daerah di mana *monkeypox* secara tradisional menjadi endemik menyoroti perlunya identifikasi dan diagnosis kasus yang cepat untuk mencegah penyebaran lebih lanjut.

Penelitian lain terkait metode yang akan peneliti gunakan dilakukan oleh Lestari *et al.*, (2021) mengenai peramalan penambahan pasien Covid-19 menggunakan *Support Vector Regression* (SVR). Hasil pemodelannya menunjukkan hasil yang baik dengan nilai MSE sebesar 89,6 dan nilai MAPE sebesar 13,9% yang menjadikan model tersebut memiliki kemampuan peramalan yang baik. Mengacu pada penelitian yang telah dilakukan sebelumnya, dalam penelitian ini akan dilakukan peramalan kasus harian *monkeypox* menggunakan *machine learning* dengan metode SVR. SVR merupakan hasil pengembangan dari

Support Vector Machine (SVM) dimana hasil proses prediksi diperoleh dari fungsi regresi $f(x)$ yang digunakan sebagai garis pemisah atau *hyperplane*. Kelebihan dari metode SVR adalah tingkat *error* yang rendah dalam menginterpretasi hasil dan dapat menghasilkan prediksi yang baik. Akan tetapi, metode SVR memiliki kelemahan pada saat *training* dimana apabila dilakukan dengan kumpulan data yang besar akan mengakibatkan proses *training* akan menjadi lambat (Hendayanti *et al.*, 2019).

Menanggapi kasus cacar monyet di seluruh dunia yang terus bertambah dengan cepat, penting untuk mengetahui dan menganalisis *trend* kasus cacar monyet di kemudian hari. Hasil analisis tersebut nantinya dapat digunakan sebagai acuan dalam merumuskan kebijakan yang tepat untuk pengendalian dan pencegahan kasus cacar monyet oleh pemerintah Indonesia dan dunia terutama WHO. Selain itu penelitian ini juga merupakan bentuk implementasi dari *Sustainable Development Goals* (SDGs) poin ketiga yaitu kehidupan sehat dan sejahtera (*Good health and well-being*).

METODOLOGI

1. Tinjauan Pustaka

Monkeypox

Monkeypox adalah *zoonosis* baru yang disebabkan oleh virus *monkeypox* (anggota genus *Orthopoxvirus* dari keluarga *Poxviridae*). Penyakit ini memiliki gejala yang sangat mirip dengan cacar yang telah diberantas pada tahun 1980. Meski gejalanya lebih ringan dari cacar, cacar monyet terjadi secara sporadis di beberapa wilayah Afrika, terutama Afrika bagian tengah dan barat. Penularan pada manusia terjadi melalui kontak langsung dengan orang atau hewan yang terinfeksi, atau melalui benda yang terkontaminasi virus. Penyakit ini dapat ringan dan gejalanya berlangsung 2-4 minggu, tetapi dapat berkembang menjadi penyakit yang berat bahkan kematian (kematian 3-6%) (Kementerian Kesehatan RI, 2022).

Uji Terasvirta

Uji Terasvirta adalah salah satu pengujian nonlinear yang dikembangkan berdasarkan model *neural network*. Uji ini dikembangkan melalui pendekatan ekspansi deret *Taylor* dan masuk ke dalam kelompok uji *Lagrangian Multiplier* (LM). Prosedur dalam pengujian Terasvirta dijabarkan sebagai berikut (Terasvirta *et al.*, 1993):

1. Meregresikan X_t pada $1, X_{t-1}, \dots, X_{t-p}$ dan sehingga diperoleh model linier

$$X_t = f_t + \hat{e}_t$$

Dimana $f_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p}$

Kemudian menghitung nilai jumlah kuadrat nilai residual $SSR_0 = \sum \hat{e}_t^2$

2. Meregresikan \hat{e}_t pada $1, X_{t-1}, X_{t-2}, \dots, X_{t-p}$ dan m prediktor tambahan yang berasal dari nilai koefisien suku kuadrat dan kubik yang dihasilkan dari pendekatan ekspansi deret *Taylor*. Selanjutnya, hitung jumlah kuadrat dari residualnya $SSR_1 = \sum \hat{v}_t^2$

3. Menghitung nilai statistik uji

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

dengan n adalah jumlah pengamatan.

Hipotesis yang digunakan adalah sebagai berikut

H_0 : $f(x)$ adalah fungsi linier dalam x (model linier)

H_1 : $f(x)$ adalah fungsi nonlinier dalam x (model nonlinier)

Keputusan tolak H_0 jika $F_{hit} > F_{(n-p-1-m)}$ atau p -value kurang dari taraf nyata 0,05.

Fungsi Kernel

Kernel merupakan suatu fungsi yang memetakan fitur data dari dimensi awal (rendah) ke dalam fitur baru dengan dimensi yang lebih tinggi (Agustina *et al.*, 2017). Berikut ini merupakan beberapa fungsi kernel (Hsu, Chang, & Lin, 2016):

- a. Linear

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

- b. Polinomial

$$K(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i^T \mathbf{x}_j + r)^d$$

c. *Radial Basic Function* (RBF)

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

d. Sigmoid

$$K(\mathbf{x}_i, \mathbf{x}_j) = \tanh(\gamma \mathbf{x}_i^T \mathbf{x}_j + r)$$

Pemilihan fungsi kernel yang sesuai penting untuk diperhatikan karena fungsi kernel menentukan fitur baru di lokasi *hyperplane* yang nantinya akan dicari. Fungsi kernel digunakan untuk mengimplementasikan model pada ruang berdimensi lebih tinggi (*feature space*) tanpa mendefinisikan fungsi pemetaan dari *input space* ke *feature space*, yang pada kasus nonlinear diasumsikan linear pada *feature space*.

Support Vector Regression (SVR)

Support Vector Regression (SVR) adalah adaptasi *machine learning* untuk klasifikasi berdasarkan model regresi *Support Vector Machine* (SVM) (Liu, Lee, & Chiu, 2009). SVM yang pada dasarnya suatu teknik klasifikasi dimana variabel responnya merupakan variabel ordinal, dikembangkan sebagai teknik regresi dimana variabel responnya adalah variabel numerik berupa bilangan *real* dan kontinu. SVR merupakan salah satu metode yang dapat mengatasi *overfitting* sehingga dapat bekerja dengan baik dan memiliki performa yang bagus (Suprayogi & Pardede, 2022).

Adapun detail penjelasan dari SVR menurut Cao & Wu (2016) adalah sebagai berikut: Diketahui data awal adalah \mathbf{x} dipetakan secara nonlinier ke ruang fitur dimensi yang lebih tinggi. Data *training* dinyatakan sebagai $\{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)\}$, dimana $\mathbf{x}_i = \{x_1, x_2, \dots, x_p\}^T \in R^n$, \mathbf{x}_i adalah vektor *input*, y_i adalah nilai luaran, n adalah jumlah *training set*. Rumus SVR dapat direpresentasikan seperti berikut:

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

dimana \mathbf{w} adalah vektor bobot, $\varphi(x)$ adalah fungsi yang memetakan \mathbf{x} dalam suatu dimensi, b adalah konstanta.

Untuk mengestimasi koefisien \mathbf{w} dan b pada persamaan di atas, dilakukan dengan meminimumkan fungsi resiko (*risk function*) dari persamaan berikut (Haykin, 2008).

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

atau dapat ditulis

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

di mana $\|\mathbf{w}\|$ adalah regularisasi, yaitu fungsi yang diminimalkan agar fungsinya sehalus mungkin (*flat*). Konstanta $C > 0$ adalah kompromi (*trade off*) antara ketipisan fungsi dan batas atas deviasi, yang lebih dari yang dapat ditoleransi (Smola & Schölkopf, 2004). Fungsi kerugian ε -*insensitive loss function* E_ε didefinisikan sebagai berikut

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

Fungsi regresi f adalah fungsi yang dapat mengaproksimasi semua titik (\mathbf{x}_i, y_i) dengan ketelitian ε . Semua titik yang terletak pada interval $f \pm \varepsilon$ disebut dapat titik yang dapat diterima (*feasible*). Titik di luar rentang $f \pm \varepsilon$ disebut titik yang tidak layak diterima (*infeasible*). Titik *infeasible* ini diperluas dengan variabel *slack* positif ξ_i dan ξ_i^* untuk mengatasi batasan tidak layak (*infeasible constrain*) dari masalah optimisasi. Sehingga persamaan fungsi resiko (*risk function*) berubah menjadi

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

dengan kendala

$$\begin{cases} y_i \leq \mathbf{w}^T \varphi(\mathbf{x}_i) + \varepsilon + \xi_i \\ y_i \geq \mathbf{w}^T \varphi(\mathbf{x}_i) - \varepsilon - \xi_i^* \\ \xi_i, \xi_i^* > 0 \end{cases}$$

Menurut Smola dan Schölkopf (2004), solusi dari persamaan di atas adalah dengan menggunakan koefisien *Lagrange* dengan koefisien *Lagrange* untuk setiap kendala yang ditunjukkan pada persamaan berikut.

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

dengan $\alpha_i, \alpha_i^*, \xi_i, \xi_i^* > 0$ adalah koefisien *lagrange*. *Lagrangian* menghasilkan solusi optimal parameter \mathbf{w} dalam bentuk

koefisien *lagrange* α_i dan α_i^* sebagai berikut

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

Sedangkan estimasi akhir dari b adalah $b = y_i - \mathbf{w}\varphi(\mathbf{x}_i) - \varepsilon$ untuk $0 \leq \alpha_i \leq C$ dan $b = y_i - \mathbf{w}\varphi(\mathbf{x}_i) + \varepsilon$ untuk $0 \leq \alpha_i^* \leq C$. Sehingga fungsi regresi SVR dapat dituliskan sebagai

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

Menurut Vapnik (1995), perkalian skalar (*dot product*) $\varphi(\mathbf{x}_i) \varphi(\mathbf{x})$ dapat diganti dengan fungsi kernel. Ini karena pemetaan data \mathbf{x} dari ruang input ke ruang fitur berdimensi lebih tinggi dengan fungsi φ biasanya tidak diketahui dan sulit dipahami. Dengan demikian, fungsi $f(x)$ ditulis sebagai berikut.

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

di mana a_i a_i^* dan adalah pengali *Lagrange*, $K(\mathbf{x}_i, \mathbf{x}_j)$ adalah fungsi kernel. Fungsi kernel memungkinkan model untuk diimplementasikan dalam ruang dimensi yang lebih tinggi (ruang fitur) tanpa harus mendefinisikan fungsi pemetaan dari ruang *input* ke ruang fitur. Meskipun ada beberapa pilihan fungsi kernel, fungsi kernel yang paling umum digunakan adalah fungsi *Radial Basis Function* (RBF) karena parameter yang harus ditentukan lebih sedikit dan memiliki kemampuan pemetaan nonlinear dari data pelatihan ke ruang dimensi tak terbatas (Cao & Wu, 2016). Fungsi kernel RBF didefinisikan sebagai berikut.

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

dimana γ menunjukkan *bandwidth* RBF. Dengan demikian, RBF diterapkan dalam penelitian ini sebagai fungsi kernel.

Algoritma Grid Search

Grid search merupakan pencarian parameter optimal dari SVR yang berdasarkan subset dari ruang *hyperparameter* yang berada dalam rentang nilai minimum (*lower bound*) dan nilai maksimum (*upper bound*) yang telah ditentukan (Syarif et al., 2016). *Grid search* membagi rentang parameter yang akan

dioptimalkan ke dalam *grid* dan iterasi melalui semua titik untuk menemukan parameter optimal. Algoritma *grid search* mencari parameter optimal SVR *cross validation*. Tujuannya adalah untuk mengidentifikasi kombinasi *hyperparameter* yang optimal untuk memprediksi data *testing* secara akurat. Oleh karena itu disarankan untuk mencoba beberapa variasi pasangan parameter pada *hyperplane* SVR (Yasin et al., 2014).

Butuh waktu lama untuk menyelesaikan pencarian *hyperparameter* yang optimal melalui metode *grid search*. Oleh karena itu, disarankan untuk melakukan pencarian *hyperparameter* menggunakan metode *grid search* dalam dua langkah, yaitu *loose grid* dan *finer grid* (Purnama & Hendarsin, 2020). *Loose grid* adalah langkah parameter optimal yang dipilih dari pangkat bilangan bulat. *Finer grid* adalah langkah selanjutnya dari *loose grid*, di mana pencarian parameter optimal didasarkan pada nilai persekitaran parameter yang dihasilkan pada langkah *loose grid*. Berdasarkan hal tersebut, penelitian ini menggunakan pencarian *hyperparameter* optimal dengan metode *grid search* dalam dua langkah, yaitu. *loose grid* dan *finer grid*. Parameter optimal adalah parameter yang memiliki keakuratan terbaik dengan nilai *error* terkecil.

Root Mean Square Error (RMSE)

Salah satu ukuran untuk mengukur tingkat akurasi peramalan adalah dengan menggunakan *Root Mean Square Error* (RMSE). RMSE adalah standar deviasi dari residual (kesalahan prediksi). RMSE adalah ukuran seberapa tersebar dari residual. Dengan kata lain, RMSE menyatakan seberapa terkonsentrasi data di sekitar garis prediksi. RMSE umumnya digunakan dalam klimatologi, peramalan, dan analisis regresi untuk memverifikasi hasil eksperimen. RMSE dapat dihitung dengan rumus berikut:

$$RMSE = \sqrt{\frac{\sum_{t=1}^n (\hat{y}_t - y_t)^2}{n}}$$

Dimana y_t merupakan data aktual pada periode ke- t , \hat{y}_t merupakan data prediksi

pada periode ke- t , dan n merupakan banyak periode waktu.

2. Data dan Sumber Data

Data yang digunakan pada penelitian ini adalah data sekunder diambil dari *website Our World In Data (OWID)* yang merupakan penerbitan daring yang menyajikan kumpulan data dengan lisensi terbuka. Data yang diambil merupakan data kuantitatif jumlah kasus terjangkit virus *monkeypox* dari seluruh dunia periode 29 Mei 2022 sampai dengan 20 Oktober 2022.

3. Variabel Penelitian

Variabel yang digunakan dalam penelitian ini adalah jumlah kasus harian *monkeypox*. Variabel pada penelitian tersaji pada Tabel 1 berikut:

Tabel 1. Variabel Penelitian

Variabel	Keterangan
Y_t	Jumlah Kasus Harian <i>Monkeypox</i>

Data yang digunakan pada penelitian ini dibagi menjadi dua *dataset* yaitu data *training* untuk membentuk model SVR dan data *testing* sebagai evaluasi model SVR yang terbentuk dari data *training* yang nantinya akan dipakai untuk melihat akurasi peramalan. Pembagian *dataset* didasarkan pada kriteria *Woschnagg & Cipan (2004)* dimana data *training* dan data *testing* memiliki perbandingan rasio 8:2 dari total data.

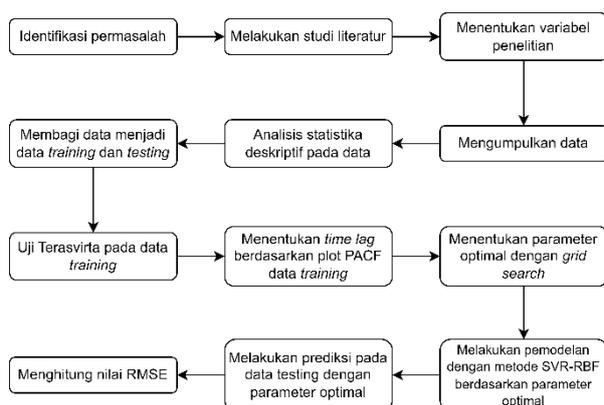
4. Prosedur Analisis

Langkah-langkah analisis yang dilakukan pada penelitian ini adalah sebagai berikut:

1. Mengidentifikasi masalah, dalam penelitian ini adalah bagaimana memprediksi secara akurat kasus harian *monkeypox* di seluruh dunia.
2. Melakukan studi literatur atau kepastakaan. Studi literatur terkait konsep peramalan dan prediksi menggunakan SVR.
3. Menentukan variabel penelitian, yaitu data jumlah kasus harian penyakit *monkeypox* di seluruh dunia.
4. Mengumpulkan data jumlah kasus harian penyakit *monkeypox* di seluruh dunia periode 29 Mei 2022 sampai dengan 20 Oktober 2022.

5. Melakukan analisis statistika deskriptif pada data.
6. Membagi data menjadi 2 *dataset* yaitu data *training* dan data *testing* dengan rasio 8:2.
7. Melakukan pengujian linearitas menggunakan uji Terasvirta pada data *training*.
8. Melakukan pemodelan pada data *training* dengan metode SVR menggunakan *software R*. Langkah-langkah pemodelan data dengan metode SVR adalah sebagai berikut:
 - a. Penentuan variabel *input* pada model SVR berdasarkan plot PACF data *training*.
 - b. Penentuan model terbaik dengan mencari parameter optimal dengan algoritma *grid search*.
 - c. Melakukan pemodelan data dengan metode SVR fungsi *Kernel Radial Basis Function (RBF)* berdasarkan parameter optimal.
 - d. Memprediksi data *testing* menggunakan model terbaik pada langkah b.
 - e. Menghitung nilai RMSE pada hasil pemodelan dan hasil prediksi pada model SVR.

Langkah-langkah analisis di atas dapat digambarkan dalam diagram alir penelitian pada Gambar 1 berikut.



Gambar 1. Diagram Alir Penelitian

HASIL DAN PEMBAHASAN

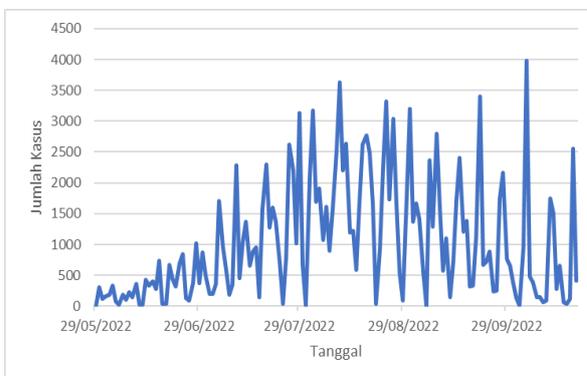
Statistika Deskriptif

Berikut ini ditampilkan nilai statistika deskriptif dari data yang digunakan.

Tabel 2. Statistika Deskriptif Jumlah Kasus Harian Virus *Monkeypox*

Variabel	Mean	StDev	Max	Min	n
Y_t	1034,6	951,1	3986		2

Berdasarkan Tabel 2, dengan menggunakan jumlah kasus harian virus *monkeypox* di seluruh dunia pada tanggal 29 Mei 2022 hingga 20 Oktober 2022, diperoleh rata-rata kasus harian *monkeypox* sebesar 1034,6 atau 1034 kasus. Nilai simpangan baku hasil statistika deskriptif menunjukkan angka yang cukup besar yaitu 951,1 dengan nilai maksimum kasus harian *monkeypox* sebesar 3986.



Gambar 2. Jumlah Kasus Harian *Monkeypox* Dunia

Berdasarkan Gambar 2 terlihat mulai terjadi kenaikan secara signifikan jumlah kasus terjangkit *monkeypox* di seluruh dunia mulai dari akhir bulan Mei. Menurut informasi tersebut, berdasarkan laporan yang diterbitkan oleh *Reuters* tanggal 25 Mei 2022, lebih dari 20 negara non-endemik melaporkan kasus *monkeypox* dan telah mengonfirmasi lebih dari 100 kasus. Kemunculan kasus baru tersebut paling banyak terjadi di Eropa. Kemudian, tanggal 20 Juli 2022, peningkatan kasus sebanyak lebih dari 14.000 kasus di seluruh dunia dilaporkan *World Health Organization* (WHO). Sejak peningkatan ini, pada 23 Juli 2022, WHO menyatakan cacar monyet atau *monkeypox* sebagai darurat kesehatan global.

Dalam pemodelan dan prediksi yang melibatkan data *time series*, hal yang perlu diperhatikan adalah pola datanya. Dalam

Gambar 2 juga dapat ditunjukkan bahwa nilai simpangan baku yang cukup besar digambarkan melalui garis yang fluktuatif dengan *trend* yang cenderung naik tiap harinya.

Pengujian Linearitas Data dengan Uji Terasvirta

Sebelum dilakukan analisis lebih lanjut, data jumlah kasus harian *monkeypox* akan diuji linearitasnya, untuk melihat apakah data cocok dimodelkan menggunakan metode nonlinear seperti SVR. Hasil pengujian linearitas pada data menggunakan uji Terasvirta disajikan pada Tabel 3.

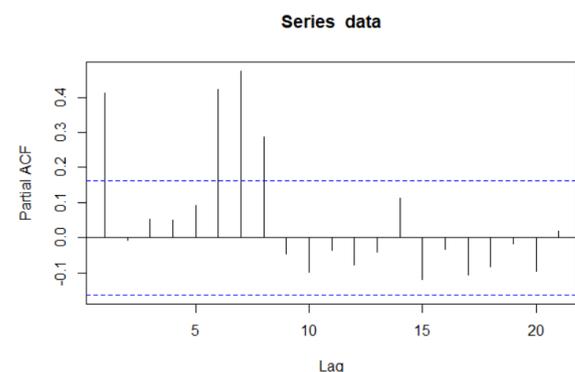
Tabel 3. Uji Terasvirta pada Data Jumlah Kasus Harian *Monkeypox*

Variabel	<i>p-value</i>
Y_t	0,004618

Berdasarkan Tabel 3, ditunjukkan bahwa nilai *p-value* hasil uji terasvirta sebesar 0,004618 lebih kecil dari taraf signifikan (0,05) sehingga dapat disimpulkan bahwa data jumlah kasus harian *monkeypox* mengandung pola nonlinear, sehingga data perlu dianalisis menggunakan model nonlinear untuk mendapatkan prediksi model yang lebih baik.

Penentuan Variabel Input Berdasarkan Plot PACF

Langkah awal sebelum melakukan pemodelan menggunakan metode SVR adalah menentukan *lag* yang signifikan dari data dimana nantinya dipakai sebagai variabel *input* untuk model. *Lag* signifikan ditentukan berdasarkan plot PACF dari data. Plot PACF dari data disajikan pada Gambar 3.



Gambar 3. Plot PACF Data Jumlah Kasus Harian *Monkeypox* Dunia

Berdasarkan Gambar 3 tampak bahwa *lag* yang signifikan adalah *lag* 1, *lag* 6, *lag* 7, dan *lag* 8. Dengan demikian, variabel input pada model SVR pada penelitian ini adalah *lag* 1, *lag* 6, *lag* 7, dan *lag* 8.

Pemodelan Data Kasus Harian *Monkeypox* Dunia Menggunakan SVR

Saat pembentukan model SVR, perlu ditentukan parameter optimal agar model mampu memprediksi secara akurat. Pencarian parameter optimal dalam penelitian ini menggunakan algoritma *grid search* dengan parameter yang dicari adalah *C*, γ dan ϵ . Penentuan nilai optimal dari parameter pada menggunakan *grid search* pada penelitian ini dilakukan melalui dua langkah yaitu langkah *loose grid* dan langkah *finer grid*.

Pada tahap awal proses pencarian parameter optimal pada algoritma *grid search* perlu ditentukan *range* nilai parameternya. Pada langkah *loose grid*, digunakan nilai pangkat bilangan bulat untuk parameter *C* dan γ . Sedangkan nilai ϵ memiliki rentang nilai dari 0 hingga 1. Tabel 4 menunjukkan rentang nilai parameter yang diimplementasikan dalam pencarian parameter optimal pada langkah *loose grid*.

Tabel 4. Rentang Nilai Parameter Tahapan *Loose Grid* Metode *Grid Search*

Parameter	Rentang Nilai
<i>C</i>	$2^{-3}, 2^{-2}, \dots, 2^5, 2^6$
γ	$2, 2^2, \dots, 2^9, 2^{10}$
ϵ	0,01, 0,02, ..., 0,1

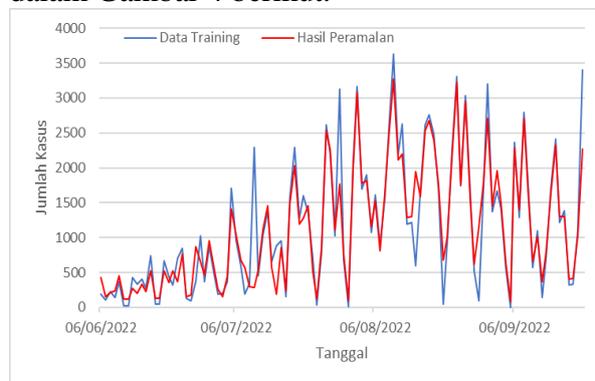
Berdasarkan Tabel 4, diperoleh nilai parameter optimal dari langkah *loose grid* untuk masing-masing parameter adalah *C* = 2, γ = 2, dan ϵ = 0,09. Pada langkah selanjutnya yaitu tahap *finer grid* akan digunakan nilai persekitaran *C* dan γ pada langkah *loose grid* untuk mencari parameter optimalnya. Rentang nilai parameter yang diimplementasikan pada langkah *finer grid* ditunjukkan pada Tabel 5.

Tabel 5. Rentang Nilai Parameter Tahapan *Finer Grid* Metode *Grid Search*

Parameter	Rentang Nilai
<i>C</i>	$2^{0,2}, 2^{0,4}, \dots, 2^{1,6}, 2^{1,8}$

γ	$2^{0,2}, 2^{0,4}, \dots, 2^{1,6}, 2^{1,8}$
ϵ	0,09

Berdasarkan hasil pencarian parameter optimal pada langkah *finer grid* diperoleh parameter optimalnya dari masing-masing parameter adalah *C* = $2^{0,6}$, γ = $2^{0,2}$, dan ϵ = 0,09. Tahapan berikutnya setelah mendapatkan model SVR terbaik adalah melihat seberapa akurat model SVR untuk peramalan data kasus harian *monkeypox* melalui nilai RMSE dan plot *time series* antara data *training* dengan hasil peramalan. Plot *time series* antara data *training* dengan hasil peramalan disajikan dalam Gambar 4 berikut.

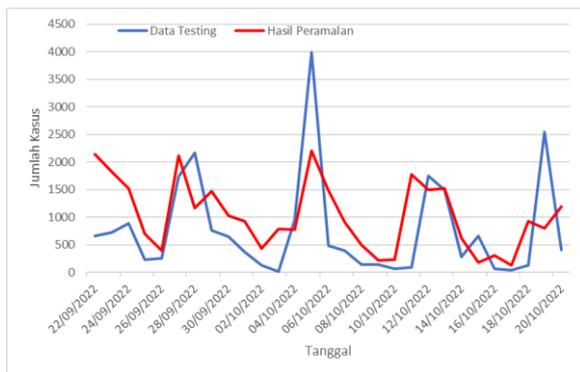


Gambar 4. Plot *Time Series* Hasil Peramalan dengan Data *Training*

Berdasarkan Gambar 4, terlihat bahwa model mampu memprediksi pada data *training* dengan cukup akurat. Hal tersebut terlihat dari hasil peramalan yang nilainya mendekati data aktualnya dan kesesuaian *trend* antara hasil peramalan dengan data aktualnya. Selain itu, diperoleh nilai RMSE pada data *training* sebesar 352,3. Sehingga, jika melihat hasil plot *time series* dan nilai RMSE pada *training*, dapat disimpulkan bahwa model SVR dengan parameter optimal mampu memprediksi data *training* dengan cukup baik.

Peramalan Data Kasus Harian *Monkeypox* Dunia Menggunakan SVR

Model SVR yang telah diperoleh kemudian diimplementasikan dalam meramalkan pada data *testing* dengan hasil plot *time series*-nya disajikan pada Gambar 5 berikut.



Gambar 5. Plot *Time Series* Hasil Peramalan dengan Data *Testing*

Berdasarkan Gambar 5 hasil peramalan nilainya mendekati nilai data *testing*-nya dengan nilai RMSE pada hasil peramalan data *testing*-nya diperoleh sebesar 809,7. Selain itu, hasil peramalan jika dilihat secara *trend*, memiliki *trend* yang sama dengan data *testing*-nya. Sehingga berdasarkan kedekatan nilai, kesamaan *trend*, dan nilai RMSE, maka model SVR dengan parameter optimal bisa dikatakan cukup baik dalam melakukan peramalan pada data *testing*.

KESIMPULAN

Berdasarkan hasil analisis diperoleh kesimpulan bahwa hasil peramalan kasus harian *monkeypox* dunia menggunakan metode SVR, baik pada data *training* maupun *testing*-nya, nilai prediksinya mendekati nilai data aktualnya. Selain itu, hasil peramalan jika dilihat secara *trend*, memiliki *trend* yang sama dengan data aktualnya. Sehingga berdasarkan kedekatan nilai dan *trend* maka model SVR bisa dikatakan cukup baik dalam melakukan peramalan. Selain itu, diperoleh nilai RMSE pada data *training* dan *testing* masing-masing sebesar 352,3 dan 809,7. Saran yang bisa diberikan untuk penelitian selanjutnya adalah dimungkinkan untuk melakukan berbagai penyesuaian dalam pemilihan parameter SVR yang optimal, atau menggunakan algoritma pencarian parameter optimal yang berbeda. Selain itu, penelitian selanjutnya disarankan untuk membandingkan juga dengan metode *machine learning* lain sehingga dapat membandingkan prediksi yang lebih akurat.

DAFTAR PUSTAKA

- Agustina, S., Agoestanto, A., dan Hendikawati P. 2017. Klasifikasi Tingkat Kesejahteraan Keluarga Jawa Tengah Tahun 2015 Menggunakan Metode Regresi Logistik Ordinal dan Support Vector Machine (SVM). *UNNES Journal of Mathematics*. Vol. 6(1), 59-69.
- Cabanillas, B., Valdelvira, R., dan Akdis, C. A. (2022). Monkeypox outbreak in Europe, UK, North America, and Australia: A changing trend of a zoonotic disease. *Allergy*, Vol. 77(8), 2284–2286. <https://doi.org/10.1111/all.15393>
- Cao, G., dan Wu, L. 2016. Support vector regression with fruit fly optimization algorithm for seasonal electricity consumption forecasting. *Energy*. Vol. 115, 734-745.
- Farid, N. (2022, Agustus 19). *Kesehatan*. Tirto.id.
- Haykin, S. 2008. *Neural Networks and Learning Machines 3rd Edition*. New York: Pearson.
- Hendayanti, N. P., Suniantara, I. K., dan Nurhidayati, M. 2019. Penerapan Support Vector Regression (SVR) dalam Memprediksi Jumlah Kunjungan Wisatawan Domestik ke Bali. *Jurnal Varian*. Vol. 3(1), 43-50.
- Hsu, C.-W., Chang, C.-C., dan Lin, C.-J. (2016). *A Practical Guide to Support Vector Classification*. Diambil kembali dari <https://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf>
- Jamil, H., Tariq, W., Tahir, M. J., Mahfooz, R. S., Asghar, M. S., dan Ahmed, A. 2022. Human monkeypox expansion from the endemic to non-endemic regions: Control measures. *Annals of Medicine and Surgery*. Vol. 79. 104048.
- Kementerian Kesehatan RI. (2022). *Pedoman Pencegahan dan Pengendalian Penyakit Monkeypox*. Jakarta: Kementerian Kesehatan RI.
- Lestari, M. P., Witarsyah, D. J., dan Hamami, F. 2021. Peramalan Pertambahan Pasien Covid-19

- Menggunakan Support Vector Regression. *eProceedings of Engineering*. Vol. 8(5), 9497.
- Liu, C.-J., Lee, T.-S., dan Chiu, C.-C. 2009. Financial time series forecasting using independent component analysis and support vector regression. *Decision Support Systems*. Vol. 47(2), 115-125.
- Purnama, D. I., Hendarsin, O. P. 2020. Peramalan Jumlah Penumpang Berangkat Melalui Transportasi Udara di Sulawesi Tengah Menggunakan Support Vector Regression (SVR). *Jambura Journal of Mathematics*. Vol. 2(2), 49-59.
- Smola, A. J., dan Schölkopf, B. 2004. A tutorial on support vector regression. *Statistics and Computing*. Vol. 14(3), 199-222.
- Suhenda, D. (2022, Agustus 3). *IDI forms task force to prevent monkeypox outbreak in Indonesia*. The Jakarta Post.
- Suprayogi, D., dan Pardede, H. F. 2022. Support Vector Regression Dalam Prediksi Penurunan Jumlah Kasus Penderita Covid-19. *Journal of Information Technology and Computer Science*. Vol. 7(2), 63-70.
- Syarif, I., Prugel-Bennett, A., dan Wills, G. 2016. SVM Parameter Optimization Using Grid Search and Genetic Algorithm to Improve Classification Performance. *TELKOMNIKA*. Vol. 14(4), 1502-1509.
- Terasvirta, T., Linc, F., dan Granger, C., W., J. 1993. Power of The Neural Networks Linearity Test. *Journal of Time Series Analysis*. Vol. 14(2), 159-171.
- Thornhill, J. P., Barkati, S., Walmsley, S., Rockstroh, J., Antinori, A., Harrison, L. B., Palich, R., Nori, A., Reeves, I., Habibi, M. S., Apea, V., Boesecke, C., Vandekerckhove, L., Yakubovsky, M., Sendagorta, E., Blanco, J. L., Florence, E., Moschese, D., Maltez, F. M., Goorhuis, A., Pourcher, V., Migaud, P., Noe, S., Pintado, C., Maggi, F., Hansen, A.-B. E., Hoffmann, C., Lezama, J. I., Mussini, C., Cattelan, A. M., Makofane, K., Tan, D., Nozza, S., Nemeth, J., Klein, M. B., dan Orkin, C. M. 2022. Monkeypox Virus Infection in Humans across 16 Countries - April-June 2022. *The New England Journal of Medicine*. Vol. 387(8), 679-691.
- Vapnik, V. N. 1995. *The Nature of Statistical Learning Theory*. New York: Springer.
- Wijaya, L. D. (2022, Agustus 22). *Indonesia's First Monkeypox Patient Visited Western Europe in Early August*. Tempo.co.
- Woschnagg, E., Cipan, J. 2004. *Evaluating Forecast Accuracy*. Austria: University of Vienna.
- Yasin, H., Prahutama, A., dan Utami, T. W. 2014. Prediksi Harga Saham Menggunakan Support Vector Regression dengan Algoritma Grid Search. *Media Statistika*. Vol. 7(1), 29-35.