

PENERAPAN MODEL *LEAST SQUARE SUPPORT VECTOR MACHINE* (LSSVM) UNTUK PERAMALAN KASUS COVID-19 DI INDONESIA

Lutfi Ardining Tyas^{1*}, I Made Tirta², Yuliani Setia Dewi³

^{1,2,3} Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Jember

*e-mail: lutfiardiningtyas13@gmail.com

DOI: 10.14710/j.gauss.12.2.304-313

Article Info:

Received: 2022-07-25

Accepted: 2023-07-25

Available Online: 2023-07-28

Keywords:

lssvm; forecasting; time series; covid-19

Abstract: Forecasting is about predicting the future based on historical data and any information that might affects the forecasts. This article applies the LSSVM model to forecast Covid-19 cases in Indonesia. The purpose of this study is to find out how the LSSVM model applied and the model performances for forecasting Covid-19 cases in Indonesia, using time series data and the factors that influence it, as input features. The factor data used in this study are mobility data and daily fully vaccinated data. The research has three main objectives; first, calculate the correlation between confirmed cases data and past data (lag) of mobility and vaccination. Second, is the selection of input features based on the highest correlation coefficient value of each variable. Third, do LSSVM modeling and Covid-19 case forecasting with the optimal model. RBF kernel and grid-search algorithm with 10-fold cross-validation are used to tune model parameters. The results show that the LSSVM model provides good performance for Covid-19 forecasting and the optimal LSSVM model for forecasting Covid-19 cases in Indonesia is using time lag 14 for the mobility factor and time lag 24 for the vaccination factor.

1. PENDAHULUAN

Covid-19 pertama kali melanda kota Wuhan, China pada Desember 2019. Hal ini dipandang sebagai awal dari sebuah pandemi yang dengan cepat menyebar ke seluruh belahan dunia. Laju penularan Covid-19 di Indonesia pada pertengahan tahun 2021 mengalami peningkatan yang cukup signifikan. Pemberlakuan PPKM dan percepatan capaian vaksinasi merupakan bentuk tindakan pemerintah Indonesia untuk menekan laju penularan. Pendataan yang dilakukan oleh Satuan Tugas Penanganan Covid-19 sangat membantu untuk meneliti laju penyebaran virus di seluruh wilayah Indonesia. Berdasarkan data- data kasus yang telah dikumpulkan, maka data tersebut dapat dipelajari untuk meramalkan laju perkembangan kasus Covid-19 di Indonesia. Peramalan merupakan suatu aktivitas untuk memprediksi kejadian atau kondisi di masa depan yang didasarkan pada data historis dan informasi yang mungkin akan berdampak pada suatu peramalan tersebut (Hyndman and Athanasopoulos, 2018). Peramalan kasus Covid-19 dapat didasarkan pada data kasus sebelumnya atau data *time series* yang telah dihimpun dengan beberapa faktor yang mempengaruhinya.

Salah satu metode *machine learning* yang telah banyak digunakan dalam permasalahan peramalan *time series* ialah *Support Vector Machine* (SVM), namun SVM seringkali memakan waktu dan beban komputasi yang tinggi karena solusi *quadratic programming*. Oleh karena itu diperkenalkanlah *Least Square Support Vector Machine* (LSSVM) yang penyelesaiannya menggunakan satu set persamaan linier sehingga kompleksitas komputasinya berkurang, kecepatan penyelesaian masalah lebih cepat, dan kemampuan untuk memecahkan masalah dalam skala besar meningkat. LSSVM dapat digunakan dalam peramalan data *time series*. Peramalan dengan LSSVM, tidak memerlukan asumsi tertentu dari data (Suykens dan Vandewalle, 2002). Berdasarkan hal

tersebut, penelitian ini akan menerapkan model LSSVM untuk peramalan kasus Covid-19 di Indonesia menggunakan data *time series* kasus Covid-19 dan faktor-faktor yang mempengaruhinya dan menemukan bagaimana performa model LSSVM untuk permasalahan peramalan pada data *time series* Covid-19 di Indonesia.

2. TINJAUAN PUSTAKA

Kasus Covid-19 pertama kali terdeteksi di Indonesia pada 2 Maret 2020, setelah dua orang dilaporkan terinfeksi. Menurut para ahli, Covid-19 menular melalui *droplets* yang keluar saat orang terinfeksi bersin atau batuk (Susilo *et al.*, 2020). Hal ini menyebabkan interaksi sosial atau mobilitas antar masyarakat menjadi salah satu faktor yang menyebabkan peningkatan penularan. Pemberlakuan PPKM terbatas yang dilakukan pemerintah merupakan salah satu program untuk mengurangi mobilitas antar masyarakat.

Tindakan penanggulangan Covid-19 lain yang dilakukan oleh pemerintah Indonesia adalah vaksinasi. Secara umum, vaksinasi diberikan dengan tujuan untuk mengurangi transmisi penularan Covid-19. Sistem kekebalan tubuh yang terbentuk pada seseorang yang telah mendapatkan vaksin, akan memperlambat resiko infeksi Covid-19 (Unit Pelayanan Kesehatan Kemenkes, 2021).

Peramalan (*forecasting*) adalah tentang memprediksi kejadian atau kondisi di masa depan dengan seakurat mungkin. Aktivitas *forecasting* memerlukan semua data informasi yang tersedia, antara lain data historis dan pengetahuan akan kejadian di masa depan yang mungkin akan berdampak pada suatu peramalan tersebut (Hyndman dan Athanasopoulos, 2018). *Time series* merupakan urutan pengamatan yang diambil secara berurutan dalam waktu. Waktu yang digunakan bisa berupa mingguan, bulanan, tahunan, dan sebagainya (Box *et al.*, 2015).

Salah satu uji asumsi klasik yaitu uji normalitas. Uji ini digunakan mengetahui data berdistribusi normal atau tidak. Salah satu uji normalitas adalah uji statistik *Kolmogorov-Smirnov* (Snell dan Sprent, 1995). Hipotesis yang disusun untuk uji normalitas *Kolmogorov Smirnov* ialah:

- a. H_0 : Data berdistribusi normal.
- b. H_1 : Data tidak berdistribusi normal.

Kriteria uji untuk uji normalitas *Kolmogorov-Smirnov* ialah:

- a. Jika nilai *p-value* $\geq 0,05$ maka terima H_0 .
- b. Jika nilai *p-value* $< 0,05$ maka tolak H_0 .

Hubungan atau korelasi antar dua variabel atau lebih dapat diketahui dengan melakukan analisis korelasi. Hubungan atau korelasi antar dua variabel terbagi menjadi dua yaitu korelasi positif dan korelasi negatif. Nilai koefisien korelasi yang dinyatakan dengan r , menunjukkan kuat tidaknya hubungan antar variabel. Koefisien korelasi memiliki nilai dari rentang $-1 \leq r \leq 1$. Nilai dari r menunjukkan semakin tidak adanya korelasi jika mendekati nol (Vogt dan Johnson, 2015).

Pearson correlation digunakan untuk menguji derajat hubungan antar variabel yang berdistribusi normal. Formula dari korelasi *pearson* dinyatakan pada Persamaan (1)

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (1)$$

dengan r koefisien korelasi pearson, n jumlah sampel, x_i nilai variabel X ke- i , y_i nilai variabel Y ke- i , \bar{x} mean variabel X, dan \bar{y} mean variabel Y.

Spearman Correlation merupakan ukuran korelasi non-parametrik untuk mengukur derajat hubungan antar variabel dan tidak memerlukan asumsi distribusi apapun dari data

atau variabel. Formula untuk menghitung koefisien korelasi *Spearman* dinyatakan pada Persamaan (2):

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2-1)} \quad (2)$$

dengan ρ koefisien korelasi *Spearman*, d_i selisih antara kedua rank, dan n jumlah sampel.

Least Square Support Vector Machine (LSSVM) merupakan modifikasi dari *Support Vector Machine* (SVM). Standar penyelesaian SVM adalah diselesaikan dengan menggunakan metode *quadratic programming*, sehingga seringkali membutuhkan waktu dan beban komputasi yang lebih tinggi (Suykens dan Vandewalle, 2002).

Kelebihan LSSVM dibandingkan dengan SVM adalah LSSVM hanya membutuhkan penyelesaian dari satu set persamaan linier, dimana hal ini tentunya akan lebih mudah dan komputasinya lebih sederhana. LSSVM menggunakan batasan berupa persamaan daripada pertidaksamaan dan mengadopsi *least square linear system* sebagai *loss function*-nya. LSSVM juga memiliki konvergensi yang baik dan tingkat presisi yang tinggi (Suykens dan Vandewalle, 2002).

Berdasarkan Vu dan Zhong (2018), prosedur dari LSSVM untuk kasus peramalan data *time series* diformulasikan dengan permasalahan regresi atau estimasi fungsi. Pada LSSVM untuk estimasi fungsi, permasalahan optimasi dinyatakan pada Persamaan (3)

$$\min_{\mathbf{w}, b, e} J_P(\mathbf{w}, e) = \frac{1}{2} \mathbf{w}^T \mathbf{w} + \gamma \frac{1}{2} \sum_{k=1}^N e_k^2 \quad (3)$$

Rumusan diatas mengikuti batasan persamaan (*equality constraints*) pada Persamaan (4)

$$y_k = \mathbf{w}^T \varphi(\mathbf{x}_k) + b + e_k, \quad k = 1, 2, \dots, N \quad (4)$$

dengan \mathbf{w} adalah vektor bobot, $e_k \in R$ adalah variabel error, $\gamma \geq 0$ adalah konstanta regularisasi, $\varphi(\mathbf{x}_k)$ adalah fungsi non linier yang memetakan data pada ruang input ke ruang vektor baru dan b adalah bias.

Penyelesaian permasalahan optimasi LSSVM untuk estimasi fungsi pada Persamaan (3) dibangun formula *Lagrange* pada Persamaan (5)

$$\mathcal{L}(\mathbf{w}, b, e; \alpha) = J_P(\mathbf{w}, e) - \sum_{k=1}^N \alpha_k \{ \mathbf{w}^T \varphi(\mathbf{x}_k) + b + e_k - y_k \} \quad (5)$$

dimana α_k adalah *lagrange multipliers*. Solusi dari Persamaan (5) dapat diperoleh dengan menurunkan secara parsial terhadap \mathbf{w}, b, e_k dan α_k . Sehingga kondisi optimal pada Persamaan (6) dipenuhi.

$$\begin{cases} \frac{\partial \mathcal{L}}{\partial \mathbf{w}} = 0 & \rightarrow \mathbf{w} = \sum_{k=1}^N \alpha_k \varphi(\mathbf{x}_k) \\ \frac{\partial \mathcal{L}}{\partial b} = 0 & \rightarrow \sum_{k=1}^N \alpha_k = 0 \\ \frac{\partial \mathcal{L}}{\partial e_k} = 0 & \rightarrow \alpha_k = \gamma e_k, \quad k = 1, 2, \dots, N \\ \frac{\partial \mathcal{L}}{\partial \alpha_k} = 0 & \rightarrow \mathbf{w}^T \varphi(\mathbf{x}_k) + b + e_k - y_k = 0, \quad k = 1, 2, \dots, N \end{cases} \quad (6)$$

dengan mengeliminasi variabel \mathbf{w} dan e_k diperoleh solusi persamaan linier pada Persamaan (7),

$$\begin{bmatrix} 0 & \mathbf{1}_v^T \\ \mathbf{1}_v & \Omega + \mathbf{I}/\gamma \end{bmatrix} \begin{bmatrix} b \\ \boldsymbol{\alpha} \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{y} \end{bmatrix} \quad (7)$$

dengan mendefinisikan beberapa vektor kolom yaitu: $\mathbf{y} = [y_1; \dots; y_N]$; $\mathbf{1}_v = [1; \dots; 1]$; $\boldsymbol{\alpha} = [\alpha_1; \dots; \alpha_N]$.

Kernel trick pada yang diaplikasikan pada penyelesaian solusi sistem persamaan linier pada Persamaan (2.14) dinyatakan sebagai $\Omega_{kl} = \varphi(\mathbf{x}_k)^T \varphi(\mathbf{x}_l) = K(\mathbf{x}_k, \mathbf{x}_l)$, dengan demikian diperoleh model LSSVM untuk estimasi fungsi adalah

$$y(x) = \sum_{k=1}^N \alpha_k K(\mathbf{x}, \mathbf{x}_k) + b \quad (8)$$

dengan α_k, b adalah solusi sistem linier Persamaan (7) dan $K(\mathbf{x}, \mathbf{x}_k)$ adalah fungsi kernel.

Kernel RBF adalah kernel yang populer digunakan karena implementasinya yang mudah dan dapat memetakan training set ke dimensi yang lebih tinggi. Dengan kernel RBF maka model LSSVM memiliki dua parameter (γ, σ^2) untuk ditentukan. Pemilihan parameter (γ, σ^2) dalam penelitian ini dilakukan dengan melakukan optimalisasi *hyperparameter*, yaitu untuk memperoleh parameter yang optimal. Penentuan parameter optimal tersebut penting dilakukan untuk dapat menghasilkan model LSSVM yang memiliki akurasi yang baik. Algoritma *grid-search*, merupakan salah satu algoritma yang populer yang dilakukan untuk *tuning parameter* atau untuk memperoleh parameter yang optimal (Liashchynskyi, P., 2019). Persamaan fungsi kernel RBF disajikan pada Persamaan (9)

$$K(\mathbf{x}, \mathbf{x}_k) = \exp\left(-\frac{\|\mathbf{x}-\mathbf{x}_k\|^2}{2\sigma^2}\right) \quad (9)$$

Berdasarkan Espinoza *et al.* (2008), permasalahan peramalan dapat dimodelkan sebagai suatu model *nonlinier autoregressive exogenous* (NARX) dimana model peramalannya dapat direpresentasikan dengan memprediksi nilai masa depan *time series* y_t dengan input nilai masa lalu *time series* tersebut ($y_{t-1}, y_{t-2}, \dots, y_{t-p}$) dan nilai masa lalu *time series* x_t yang mempengaruhi y_t . Nilai masa lalu dari suatu *time series* disebut dengan *lag*. Variabel *lag* merupakan variabel yang berisi sejumlah nilai masa lalu dari variabel tersebut.

Berdasarkan penjelasan tersebut, prediktor LSSVM dapat dilatih menggunakan suatu set nilai historis *time series* ($y_{t-1}, y_{t-2}, \dots, y_{t-p}, x_{t-1}, x_{t-2}, \dots, x_{t-p}$) sebagai *input* (\mathbf{x}), untuk memprediksi nilai y_t , dengan t adalah jumlah data yang tersedia, dan p adalah panjang *lag*. Oleh karena itu, strategi modeling untuk peramalan *time series* menggunakan LSSVM dengan k adalah banyak langkah peramalan, dapat dinyatakan dengan menggunakan fungsi pada Persamaan (10), (11) dan (12) berikut

$$\hat{y}_t = f(y_{t-1}, \dots, y_{t-p}, x_{t-p}, \dots, x_{t-p}) \quad (10)$$

$$y_{t+1} = f(y_t, y_{t-1}, \dots, y_{t-p+1}, x_{t-p}, \dots, x_{t-p+1}) \quad (11)$$

$$y_{t+k} = f(y_{t+k-1}, \dots, y_{t-p+1}, x_{t-p}, \dots, x_{t-p+1}) \quad (12)$$

Metode Grid-Search digunakan untuk menentukan parameter yang optimal. Sebelum dilakukan *grid search*, terlebih dahulu dilakukan inisialisasi terhadap parameter yang akan dioptimalkan. Metode ini diimplementasikan untuk dapat men-*training* beberapa model prediksi yang disusun oleh setiap kombinasi parameter atau setiap kombinasi dari inisialisasi parameter yang diberikan. Model prediksi dikatakan sebagai model yang optimal berdasarkan *grid-search* ialah model dengan nilai *score* terbaik. Proses *training* model dengan *Grid Search* dilakukan dengan menggunakan suatu teknik validasi yaitu *cross-validation* (Liashchynskyi, P., 2019).

Salah satu teknik validasi model untuk menemukan akurasi yang baik adalah *K-Fold Cross Validation*. Teknik ini dilakukan dengan membagi dataset menjadi k *independent subset*. Salah satu dari k subset ini digunakan sebagai *validation set*, sedangkan $k - 1$

subset yang lain digunakan untuk *training set*. Model *10-fold cross validation* merupakan salah satu model *k-fold cross validation* yang sering digunakan (Berrar, 2018).

Beberapa indikator yang digunakan untuk mengevaluasi akurasi untuk validasi model antara lain adalah *Mean Absolute Error* (MAE), *Mean Square Error* (MSE), dan *Root Mean Square Error* (RMSE) dan koefisien determinasi (R^2). Indikator-indikator diatas didefinisikan pada Persamaan (10), (11), (12) dan (13):

$$MAE = \frac{1}{N} \sum_{i=1}^N |\hat{y}_i - y_i| \quad (13)$$

$$MSE = \frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2 \quad (14)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2} \quad (15)$$

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (16)$$

dengan, N adalah jumlah sampel *test*, \hat{y}_i dan y_i adalah nilai prediksi dan aktual pada waktu ke- i , dan \bar{y} adalah nilai *mean* dari nilai aktual.

3. METODE PENELITIAN

Data yang digunakan dalam penelitian ini adalah data sekunder yang meliputi data kasus terkonfirmasi harian, mobilitas harian dan vaksinasi lengkap harian di Indonesia. Data kasus terkonfirmasi Covid-19 diperoleh dari *website The Humanitarian Data Exchange* yang dikelola oleh *Johns Hopkins University* (JHU CCSE) dan *website* Peta Sebaran Covid-19 yang dikelola oleh Satuan Tugas Penanganan Covid-19. Data mobilitas diperoleh dari *Covid-19 Community Mobility Report* dan data vaksinasi lengkap diperoleh dari *website The Humanitarian Data Exchange*. Data diambil dari tanggal 1 Februari 2021 sampai 28 Februari 2022.

Variabel terikat (Y) merupakan data jumlah kasus terkonfirmasi harian Covid-19 di Indonesia yang didefinisikan sebagai variabel *Daily_cases*. Variabel bebas (X) merupakan data mobilitas harian yang menunjukkan persentase tren mobilitas pada tempat-tempat yang dikategorikan menjadi 6 kategori atau variabel terdiri dari 6 variabel (*Retail and Recreation, Parks, Transit Stations, Grocery and Pharmacy, Workplaces* dan *Residential*) dan data jumlah vaksinasi Covid-19 lengkap harian yang didefinisikan sebagai variabel *Daily_fully_vaccinated*.

Langkah-langkah penelitian yang dilakukan dalam penelitian ini adalah sebagai berikut:

1. Studi literatur untuk memperoleh informasi dan materi terkait untuk penelitian.
2. Pengambilan data dan input data kedalam program *python*.
3. *Pre-processing* data meliputi pemilihan data yang diperlukan, penghapusan data yang tidak diperlukan.
4. Uji normalitas dengan uji *Kolmogorov-Smirnov*, untuk menentukan metode korelasi.
5. Perhitungan korelasi kasus Covid-19 dengan data mobilitas dan vaksinasi. Perhitungan korelasi menggunakan data dengan *time lag* atau data masa lalu masing-masing variabel mobilitas dan vaksinasi, untuk mengetahui pada *lag* ke berapa variabel mobilitas dan vaksinasi mempengaruhi kasus Covid-19.
6. Pemilihan fitur input model, didasarkan pada variabel dan panjang *lag* dengan nilai koefisien korelasi tertinggi.
7. Pemodelan LSSVM (*training model*), diawali dengan *splitting* data *training* dan data *testing*, juga inisialisasi *hyperparameter*.

8. Pengujian model untuk memperoleh model optimum.
Proses *forecasting* dengan model optimum dan evaluasi hasil peramalan.

4. HASIL DAN PEMBAHASAN

Pre-processing data terdiri dari proses *cleaning* data dengan mengambil data yang diperlukan dan melakukan transformasi data dari data kumulatif menjadi data jumlah harian. Karena LSSVM tidak memerlukan asumsi data, pada tahap ini tidak dilakukan proses uji asumsi data *time series* yang digunakan. Variabel Y yaitu *Daily_cases* merupakan data kasus terkonfirmasi harian Covid-19 di Indonesia. Variabel X yaitu data vaksinasi harian (*Daily_fully_vaccinated*) dan data mobilitas (*retail_recreation*, *grocery_pharmacy*, *parks*, *transit_stations*, *workplaces*, *residential*).

Data mobilitas menunjukkan persentase bagaimana kunjungan ke berbagai tempat berubah sejak pandemi Covid-19 terjadi, dibandingkan dengan periode *baseline*. *Baseline* merupakan periode sebelum terjadinya pandemi, yang merupakan nilai median dari mobilitas selama 5 minggu, yaitu dari 3 Januari sampai 6 Februari 2020.

Uji normalitas digunakan untuk menentukan metode perhitungan korelasi yang akan dilakukan pada penelitian ini. Hipotesis nol pada uji normalitas menginterpretasikan data berdistribusi normal. Hasil uji normalitas untuk variabel *Daily_cases* menghasilkan *p-value* $8,98 \times 10^{-23}$, sangat kecil mendekati nol. Dengan taraf signifikansi $\alpha = 0,05$, maka H_0 ditolak, sehingga variabel *Daily_cases* tidak berdistribusi normal. Berdasarkan hal ini, maka metode perhitungan korelasi yang digunakan ialah *Spearman Rank Correlation*.

Tujuan dilakukan perhitungan korelasi adalah untuk mengetahui hubungan antar variabel *Daily_cases* dengan variabel mobilitas dan variabel vaksinasi, juga melihat seberapa berpengaruhnya mobilitas dan vaksinasi terhadap jumlah kasus Covid-19.

Menurut *World Health Organization* (2020) Covid-19 memiliki masa inkubasi dalam *range* 1 sampai 14 hari. Hal ini masuk akal untuk berasumsi bahwa kasus Covid-19 berkaitan dengan mobilitas yang terjadi beberapa hari sebelumnya. Sehingga pada penelitian ini dilakukan perhitungan korelasi variabel *Daily_cases* dengan keenam variabel mobilitas menggunakan data masa lalu dari 1 sampai 14 hari sebelumnya atau *time lag* 1 sampai 14. Periode perhitungan korelasi variabel mobilitas dengan *Daily_cases* ialah 1 Maret 2021-3 Juli 2021.

Hasil perhitungan menghasilkan tiga variabel dengan koefisien korelasi tertinggi yaitu variabel *retail_recreation* pada $t - 14$ (0,487), variabel *parks* pada $t - 14$ (0,496), dan variabel *workplaces* pada $t - 12$ (0,431). Ketiga variabel lain memiliki nilai koefisien korelasi yang lebih rendah yaitu variabel *grocery_pharmacy* pada $t - 14$ (0,391), variabel *transit_stations* pada $t - 11$ (0,36), dan variabel *residential* pada $t - 1$ (0,313).

Berdasarkan Polack *et al.* (2020) vaksin akan menjadi efektif pada 3 sampai 4 minggu setelah penyuntikan. Diperlukan rentang waktu agar vaksin dapat efektif membentuk imun pada tubuh manusia sehingga dapat berdampak pada pengurangan jumlah kasus terkonfirmasi Covid-19. Sehingga pada penelitian ini juga dilakukan perhitungan korelasi variabel *Daily_cases* dengan variabel *Daily_fully_vaccinated* menggunakan data masa lalu dari 1 sampai 30 hari sebelumnya atau dengan *time lag* 1 sampai 30. Periode perhitungan korelasi variabel *Daily_fully_vaccinated* dengan *Daily_cases* ialah 1 Februari 2021-28 Februari 2022.

Hasil perhitungan korelasi menghasilkan nilai koefisien korelasi negatif. Yang menunjukkan bahwa pertambahan jumlah vaksinasi Covid-19 berakibat pada penurunan jumlah kasus Covid-19. Empat nilai koefisien korelasi negatif tertinggi dihasilkan oleh

variabel *Daily_fully_vaccinated* pada data $t - 24$ (-0,491), data $t - 25$ (-0,491), data $t - 26$ (-0,495) dan data $t - 30$ (-0,487).

Berdasarkan model LSSVM pada *time series*, prediktor atau fitur input LSSVM dapat dilatih menggunakan suatu set nilai historis *time series* ($y_{t-1}, y_{t-2}, \dots, y_{t-p}, x_{t-1}, x_{t-2}, \dots, x_{t-p}$) sebagai *input* (\mathbf{x}), untuk memprediksi nilai y_t . Dengan t adalah jumlah data yang tersedia, dan p adalah order dari variabel *lag*. Variabel *Daily_cases* sebagai variabel Y digunakan sebagai fitur input. Untuk *predictor* atau fitur input nilai masa lalu variabel X, dipilih berdasarkan nilai koefisien korelasi tertinggi digunakan sebagai fitur input pemodelan LSSVM.

Fitur input pemodelan LSSVM untuk peramalan kasus Covid-19 dengan faktor mobilitas ialah variabel *retail_recreation*, *parks* dan *workplaces* dengan masing-masing nilai koefisien korelasi tertinggi yang diperoleh adalah sebesar 0,487; 0,49; 0,431. Panjang orde *time lag* (p) yang digunakan yaitu *time lag* dengan nilai koefisien korelasi tertinggi yang diperoleh dari keenam variabel mobilitas yaitu $p_1 = 1,11,12,14$.

Fitur input pemodelan LSSVM untuk peramalan kasus Covid-19 dengan faktor vaksinasi ialah dengan mengambil empat nilai koefisien korelasi tertingginya, yaitu pada *lag* 24 (-0,491), *lag* 25 (-0,491), *lag* 26 (-0,495) dan *lag* 30 (-0,487). Sehingga orde *time lag* yang digunakan ialah $p_2 = 24,25,26,30$.

Berikut pada Tabel 1 adalah 8 model dengan masing-masing fitur inputnya.

Tabel 1. Pendefinisian model dan fitur inputnya

Model	Fitur Input	p_1	Model	Fitur Input	p_2
M1	<i>Daily_cases, retail_recreation, parks, workplaces</i>	1	V24	<i>Daily_cases, Daily_fully_vaccinated</i>	24
M11	<i>Daily_cases, retail_recreation, parks, workplaces</i>	11	V25	<i>Daily_cases, Daily_fully_vaccinated</i>	25
M12	<i>Daily_cases, retail_recreation, parks, workplaces</i>	12	V26	<i>Daily_cases, Daily_fully_vaccinated</i>	26
M14	<i>Daily_cases, retail_recreation, parks, workplaces</i>	14	V30	<i>Daily_cases, Daily_fully_vaccinated</i>	30

Data yang digunakan dalam pemodelan LSSVM ialah data dari rentang waktu 01 Februari 2021 sampai 28 Februari 2022, yaitu sebanyak 393 data untuk setiap variabelnya. *Splitting* data menjadi *training set* sebanyak 70% dan *testing set* 30%. Kernel yang digunakan ialah kernel RBF. Dalam penelitian ini digunakan *ranges* parameter $\gamma = [100,1000]$ dengan *increment* 50, dan parameter $\sigma^2 = [10,100]$ dengan *increment* 5. Pemilihan *ranges* parameter tersebut didasarkan pada *ranges* yang sering digunakan pada pemodelan LSSVM secara umum (Zhang, Li dan Zhong, 2009). Parameter tersebut di optimalisasi dengan menggunakan *grid-search algorithm* dengan *10-fold cross-validation* dalam proses *training* model.

Hasil performa model dan kombinasi parameter terbaik dari masing-masing model pada proses *training* dan pengujian model ditunjukkan pada Tabel 2 dan Tabel 3. Hasil pemodelan LSSVM untuk faktor mobilitas menghasilkan M14 dengan *time lag* 14 sebagai model optimum untuk peramalan kasus Covid-19. Hasil pemodelan LSSVM untuk faktor vaksinasi menghasilkan V24 dengan *time lag* 24 sebagai model optimum untuk peramalan.

Tabel 2. Hasil pemodelan LSSVM untuk faktor mobilitas

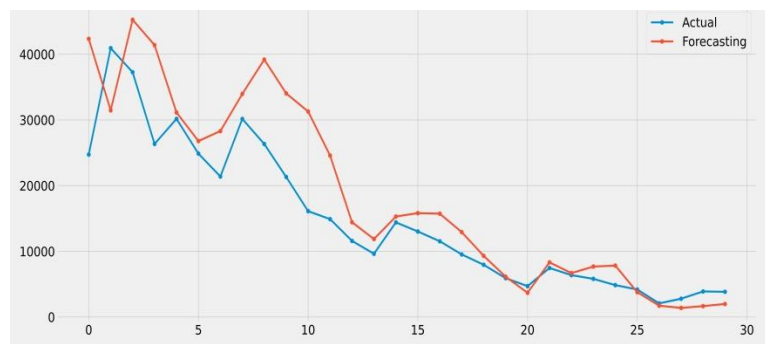
Model	Training					Testing				
	γ	σ^2	MAE ($\times 10^3$)	MSE ($\times 10^6$)	RMSE ($\times 10^3$)	R^2	MAE ($\times 10^3$)	MSE ($\times 10^6$)	RMSE ($\times 10^3$)	R^2

M1	950	10	1,588	7,837	2,799	0,947	2,089	21,867	4,676	0,941
M11	950	10	1,181	3,068	1,751	0,979	2,099	14,112	3,756	0,962
M12	950	10	1,141	2,947	1,716	0,98	2,103	13,612	3,689	0,963
M14	950	10	1,087	2,689	1,639	0,982	2,120	13,432	3,665	0,964

Tabel 3. Hasil pemodelan LSSVM untuk faktor vaksinasi

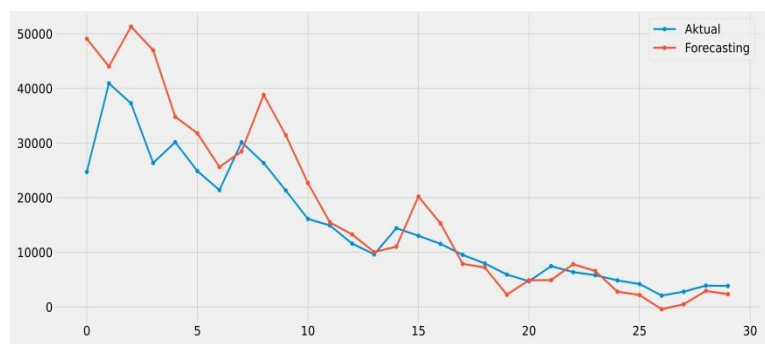
Model	Training						Testing			
	γ	σ^2	MAE ($\times 10^3$)	MSE ($\times 10^6$)	RMSE ($\times 10^3$)	R^2	MAE ($\times 10^3$)	MSE ($\times 10^6$)	RMSE ($\times 10^3$)	R^2
V24	950	20	1,376	5,691	2,385	0,967	2,355	17,741	4,212	0,952
V25	700	20	1,424	6,106	2,471	0,965	2,414	19,154	4,321	0,95
V26	550	20	1,465	6,453	2,540	0,963	2,465	20,231	4,498	0,946
V30	800	25	1,472	6,620	2,573	0,962	2,571	21,450	4,631	0,942

Model optimal yang diperoleh ialah M14 untuk peramalan dengan faktor mobilitas dan V24 untuk peramalan dengan faktor vaksinasi. Kedua model tersebut masing-masing digunakan untuk melakukan peramalan kasus terkonfirmasi Covid-19 pada 30 hari kedepan yaitu pada bulan Maret 2022. Fitur input data historis yang digunakan untuk peramalan ialah menggunakan data aktual pada bulan sebelumnya yaitu Februari dan data aktual bulan Maret 2022.



Gambar 1. Hasil peramalan model optimum faktor mobilitas

Nilai *error* hasil peramalan yang diperoleh oleh M14 yaitu $MAE = 4,843 \times 10^3$, $MSE = 49,493 \times 10^6$, dan $RMSE = 7,035 \times 10^3$. Nilai akurasi R^2 yang diperoleh 0,573.



Gambar 2. Hasil peramalan model optimum faktor vaksinasi

Nilai error hasil peramalan yang diperoleh oleh V24 yaitu $MAE = 4,934 \times 10^3$, $MSE = 58,202 \times 10^6$, dan $RMSE = 7,629 \times 10^3$. Nilai akurasi R^2 yang diperoleh 0,5.

Hasil pada proses peramalan pada bulan Maret 2020 menghasilkan hasil yang cukup baik. Dengan tren hasil prediksi mengikuti tren data aktualnya yaitu menurun dan fluktuatif. Nilai akurasi R^2 yang diperoleh dari kedua model juga tidak jauh berbeda. Berdasarkan nilai error dan akurasi R^2 , yang diperoleh pada peramalan model dengan data *testing* dan peramalan model pada data baru, diperoleh bahwa model peramalan kasus Covid-19 di Indonesia dengan faktor mobilitas lebih baik dibandingkan dengan peramalan dengan faktor vaksinasi. Fitur input data historis kasus terkonfirmasi harian dan data mobilitas di *Retail and Recreation*, *Parks* dan *Workplaces* pada 1 sampai 14 hari sebelumnya, memiliki performa yang baik untuk meramalkan kasus terkonfirmasi Covid-19 di Indonesia.

5. KESIMPULAN

Berdasarkan analisis yang telah dilakukan, maka dapat diperoleh kesimpulan yaitu, model LSSVM optimal untuk peramalan kasus Covid-19 di Indonesia pada 1 Februari 2021-28 Februari 2022 yaitu model M14 dan V24. Model M14 dengan fitur input data 1 sampai 14 hari sebelumnya dari variabel *Daily_cases*, *retail_recreation*, *parks* dan *workplaces*. Model V24 dengan fitur input data 1 sampai 24 hari sebelumnya dari variabel *Daily_cases* dan *Daily_fully_vaccinated*.

Kedua model optimal menghasilkan performa yang cukup baik saat digunakan untuk peramalan bulan Maret 2022 dengan nilai akurasi $R^2 = 0,573$ untuk M14 dan $R^2 = 0,5$ untuk V24. Model peramalan kasus Covid-19 di Indonesia dengan faktor mobilitas M14, menghasilkan performa model yang lebih baik dibandingkan dengan model peramalan dengan faktor vaksinasi V24. Hal ini diperoleh dari nilai error yang lebih kecil dan nilai akurasi R^2 yang lebih tinggi dari model peramalan dengan faktor vaksinasi.

DAFTAR PUSTAKA

- Berrar, D. (2018) 'Cross-validation', *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics*, 1–3, pp. 542–545. doi:10.1016/B978-0-12-809633-8.20349-X.
- Box, G.E. et al. (2015) *Time series analysis: forecasting and control*. John Wiley & Sons.
- Espinoza, M. et al. (2008) 'Time Series Prediction using LS-SVMs', *European Symposium on Time Series Prediction*, pp. 159--168.
- Google LCC. 2022. *COVID-19 Community Mobility Reports*, 2022. <https://www.google.com/covid19/mobility/>. Diakses : 10 Maret 2022.
- Hyndman, R.J. and Athanasopoulos, G. (2018) *Forecasting: principles and practice*. OTexts.
- Johns Hopkins University. *Novel Coronavirus (COVID-19) Cases Data*. <https://data.humdata.org/dataset/novel-coronavirus-2019-ncov-cases>. Diakses:10 Maret 2022.
- Liashchynskiy, Petro and Liashchynskiy, Pavlo (2019) 'Grid Search, Random Search, Genetic Algorithm: A Big Comparison for NAS', *ArXiv [Preprint]*, (December 2019).

- Polack, F.P. *et al.* (2020) ‘Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine’, *New England Journal of Medicine*, 383(27), pp. 2603–2615. doi:10.1056/nejmoa2034577.
- Snell, J. and Sprent, P. (1995) *Applied Nonparametric Statistical Methods.*, *Journal of the Royal Statistical Society. Series A (Statistics in Society)*. doi:10.2307/2983315.
- Susilo, A. *et al.* (2020) ‘Coronavirus Disease 2019: Tinjauan Literatur Terkini’, *Jurnal Penyakit Dalam Indonesia*, 7(1), p. 45. doi:10.7454/jpdi.v7i1.415.
- Suykens, J.A.K. and Vandewalle, J. (2002) *Least squares support vector machines*. World Scientific Publishing Co. Pte. Ltd.
- The Humanitarian Data Exchange. *Coronavirus (COVID-19) Vaccinations*. <https://data.humdata.org/dataset/covid-19-vaccinations>. Diakses:10 April 2022.
- Unit Pelayanan Kesehatan Kemenkes. (2021) *4 Manfaat Vaksin Covid-19 yang Wajib Diketahui*. <https://upk.kemkes.go.id/new/4-manfaat-vaksin-covid-19-yang-wajib-diketahui>. Diakses : 19 April 2022.
- Vogt, W. and Johnson, R. (2015) *Correlation and Regression Analysis, The Statistical, Economic and Social Research and Training Centre for Islamic Countries (SESRIC)*. doi:10.4135/9781446286104.
- Vu, G.X.. and Zhong, Z.. (2018) ‘Forecasting Air Passengers of Changi Airport Based on Seasonal Decomposition and an LSSVM Model’, *Review of Information Engineering and Applications*, 5(1), pp. 12–30. doi:10.18488/journal.79.2018.51.12.30.
- World Health Organization (2020) ‘COVID-19 Situation Report 73’, *World Health Organization*, 2019(March), p. 2633. doi:10.1056/NEJMoa2001316.4.
- Zhang, W., Li, C. and Zhong, B. (2009) ‘LSSVM parameters optimizing and non-linear system prediction based on cross validation’, *5th International Conference on Natural Computation, ICNC 2009*, 1(3), pp. 531–535. doi:10.1109/ICNC.2009.26.