

Imputasi Data Temperatur Maksimum Menggunakan Metode Support Vector Regression

Isa Kholifatus Sukhna ^{(1)*}, Brina Miftahurrohmah ⁽²⁾, Catur Wulandari ⁽³⁾, Putri Amelia ⁽⁴⁾
Departemen Sistem Informasi, Universitas Internasional Semen Indonesia, Gresik, Indonesia
e-mail : isa.sukhna20@student.uisi.ac.id,
{brina.miftahurrohmah,catur.wulandari,putri.amelia}@uisi.ac.id.

* Penulis korespondensi.

Artikel ini diajukan 18 Mei 2024, direvisi 22 Agustus 2024, diterima 22 Agustus 2024, dan dipublikasikan 31 Mei 2025.

Abstract

Temperature is a crucial element affecting various aspects, from agriculture to natural disasters. Temperature data imputation is also important because, in some cases, temperature data is not always complete. This study aims to predict missing temperature data in the East Nusa Tenggara (NTT) region using the Support Vector Regression (SVR) method. The data used comes from six BMKG observation stations in NTT and ERA-5 Reanalysis data. The choice of the SVR method is based on its ability to handle data with complex structures. Modeling is conducted separately for each station using the Radial Basis Function (RBF) kernel. Model evaluation employs the metrics Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Coefficient of Determination (R^2), presenting the evaluation results with low error. The results show that among several parameter tests, the parameter ranges $[C = 1, 5, 10, 15]$, $[\epsilon = 0, 1, 0, 3, 0, 6, 0, 9]$, and $[\gamma = 1, 5, 10, 15]$ in the SVR method are the best parameter ranges across all stations. The prediction graphs display different temperature fluctuation patterns at each station. This study contributes to enhancing the availability of accurate climate data, supporting sustainable decision-making in the NTT region.

Keywords: Temperature, SVR, NTT, Reanalysis ERA-5, RBF

Abstrak

Temperatur merupakan elemen penting yang mempengaruhi berbagai aspek mulai dari pertanian hingga bencana alam. Imputasi data temperatur juga menjadi penting karena dalam beberapa kasus data temperatur tidak selalu lengkap. Penelitian ini bertujuan untuk memprediksi data temperatur yang hilang di wilayah Nusa Tenggara Timur (NTT) menggunakan metode Support Vector Regression (SVR). Data yang digunakan berasal dari 6 stasiun pengamatan BMKG di NTT dan data Reanalysis ERA-5. Pemilihan metode SVR didasarkan pada kemampuannya dalam menangani data dengan struktur kompleks. Pemodelan dilakukan secara terpisah pada setiap stasiun dengan menggunakan fungsi kernel Radial Basis Function (RBF). Evaluasi model menggunakan metrik Root Mean Square Error (RMSE), Mean Absolute Error (MAE), dan Coefficient of Determination (R^2) dengan menampilkan hasil evaluasi dengan error yang rendah. Hasil penelitian menunjukkan bahwa dari beberapa pengujian parameter, rentan parameter $[C = 1, 5, 10, 15]$, $[\epsilon = 0, 1, 0, 3, 0, 6, 0, 9]$, $[\gamma = 1, 5, 10, 15]$ dalam metode SVR merupakan rentan parameter terbaik pada semua stasiun. Grafik prediksi menampilkan pola fluktuasi temperatur yang berbeda-beda pada setiap stasiun. Penelitian ini berkontribusi dalam meningkatkan ketersediaan data iklim yang akurat untuk mendukung pengambilan keputusan yang berkelanjutan di wilayah NTT.

Kata Kunci: Temperatur, SVR, NTT, Reanalysis ERA-5, RBF

1. PENDAHULUAN

Iklim adalah salah satu komponen dalam ekosistem yang tidak terpisahkan dari kehidupan makhluk hidup. Iklim berkaitan erat dengan kondisi cuaca rata-rata di suatu wilayah tertentu selama periode waktu yang panjang (Malihah, 2022). Perubahan iklim memiliki dampak yang signifikan pada berbagai sektor, dan salah satunya adalah sektor pertanian. Di Indonesia,



pertanian memainkan peran penting dengan banyaknya penduduk yang bekerja di sektor tersebut. Sektor pertanian sangat rentan terhadap perubahan iklim karena perubahan tersebut dapat memengaruhi berbagai aspek seperti pola tanam, waktu tanam, produksi, dan kualitas hasil. Produksi pertanian dipengaruhi oleh faktor-faktor seperti peningkatan suhu, perubahan pola curah hujan, penguapan air, dan kelembapan tanah, yang semuanya berpotensi memengaruhi produktivitas (Harvian & Yuhan, 2021). Perubahan iklim ekstrem juga dapat berpotensi terjadinya bencana alam. Menurut laporan Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) Stasiun Meteorologi El Tari Kupang, intensitas musim hujan di wilayah Nusa Tenggara Timur (NTT) hingga tahun 2022 sangat rendah dibandingkan dengan musim kemarau yang berlangsung selama 9 bulan. Selain itu pada tahun 2021, Kabupaten Alor merupakan salah satu daerah yang berpotensi mengalami kebakaran hutan. Berdasarkan data BPBD Provinsi Nusa Tenggara Timur yang dilaporkan oleh Kompas, dari Juni hingga Agustus 2021 terdapat 165 titik panas yang tersebar hampir merata di 22 kabupaten/kota di NTT. Titik panas ini diperkirakan akan terus menyebar ke wilayah timur mengikuti arah angin dan musim kemarau di wilayah NTT (Natalia & Firdaus, 2021).

Nusa Tenggara Timur (NTT) merupakan sebuah provinsi di Indonesia yang mencakup bagian timur Kepulauan Nusa Tenggara. Provinsi ini berbatasan dengan Laut Flores di utara, Samudera Hindia di selatan, Timor Leste di timur, dan Provinsi Nusa Tenggara Barat di barat. Ibu kota NTT adalah Kupang, yang terletak di antara $8^{\circ} - 12^{\circ}$ Lintang Selatan dan $118^{\circ} - 125^{\circ}$ Bujur Timur, serta terdiri dari 22 kabupaten/kota (BPK RI Perwakilan Provinsi Nusa Tenggara Timur, 2024). Penentuan Wilayah NTT sebagai tempat pelaksanaan penelitian didasari oleh karakteristik unik iklimnya yang menonjol dan fluktuasi temperatur udara yang sangat bervariasi sepanjang tahun. Faktor-faktor ini menjadikan NTT sebagai lokasi yang relevan dan signifikan untuk penelitian yang menitikberatkan pada prediksi temperatur. Selain itu, tersedia stasiun cuaca yang mewakili wilayah tersebut, yang dapat diandalkan sebagai sumber data historis temperatur.

Berbagai cara telah dikembangkan untuk menangani data yang tidak lengkap (*missing data*). Secara umum, penanganan ini dapat dibagi menjadi dua kategori utama yaitu metode imputasi dan metode amputasi. Dalam banyak situasi metode amputasi sering tidak disarankan dan metode imputasi cenderung lebih efektif. Beberapa penelitian juga menunjukkan bahwa penggunaan metode imputasi untuk mengatasi data yang hilang atau tidak lengkap dapat meningkatkan akurasi klasifikasi dibandingkan dengan tidak menggunakan metode imputasi sama sekali (Agwil et al., 2024). Terdapat beberapa penelitian terdahulu misalnya dari Jabir et al. (2023) yang melakukan imputasi menggunakan metode *mean imputation* di mana untuk mencari data yang hilang yaitu dengan melakukan rata-rata pada masing-masing variabel. Pada penelitian Laia (2023) melakukan imputasi menggunakan metode K-Nearest Neighbor *imputation* yaitu dengan melakukan pendekatan yang digunakan untuk melakukan klasifikasi pada data terdekat. Penelitian ketiga dari Sefidian & Daneshpour (2019) menggunakan kombinasi Teori Grey dan menggunakan model regresi Support Vector Regression (SVR) untuk melakukan imputasi. Dari beberapa penelitian tersebut menghasilkan penelitian dengan judul “Imputasi Data Temperatur Maksimum Pada Wilayah NTT Menggunakan Metode Support Vector Regression” karena tidak ada yang melakukan penelitian berupa imputasi data hilang pada Wilayah Nusa Tenggara Timur menggunakan metode Support Vector Regression.

Pemilihan perhitungan temperatur sebagai fokus penelitian disebabkan oleh kepentingan temperatur sebagai salah satu parameter iklim yang krusial dan mempunyai dampak besar di sejumlah sektor, termasuk pertanian, ketahanan pangan hingga bencana alam. Terdapatnya data *missing* merupakan hal yang sering terjadi, hal ini disebabkan oleh berbagai alasan seperti kerusakan alat, perhitungan yang tidak akurat, data yang tidak tercatat, cuaca yang tidak mendukung, serta berbagai faktor teknis lainnya (Prasetya et al., 2023; Septiansari et al., 2021). Ketersediaan informasi komponen cuaca dan iklim seperti temperatur sangat diperlukan baik dari segi data statistik maupun data spasialnya (Pratikno & Rahardjo, 2021). Ketersediaan data yang akurat sangat diperlukan untuk memahami dinamika iklim, melaksanakan peramalan cuaca yang tepat, serta untuk mendukung pengambilan keputusan yang lebih terinformasi di beragam aspek kehidupan. Dengan demikian, penelitian ini bertujuan untuk mengatasi kendala data temperatur



yang hilang di wilayah NTT dengan memanfaatkan metode Support Vector Regression untuk memprediksi data yang hilang. Diharapkan hasil penelitian akan memberikan kontribusi yang signifikan dalam memperbaiki ketersediaan data iklim yang akurat dan mendukung beragam upaya perencanaan dan pengambilan keputusan yang berkelanjutan di wilayah Nusa Tenggara Timur.

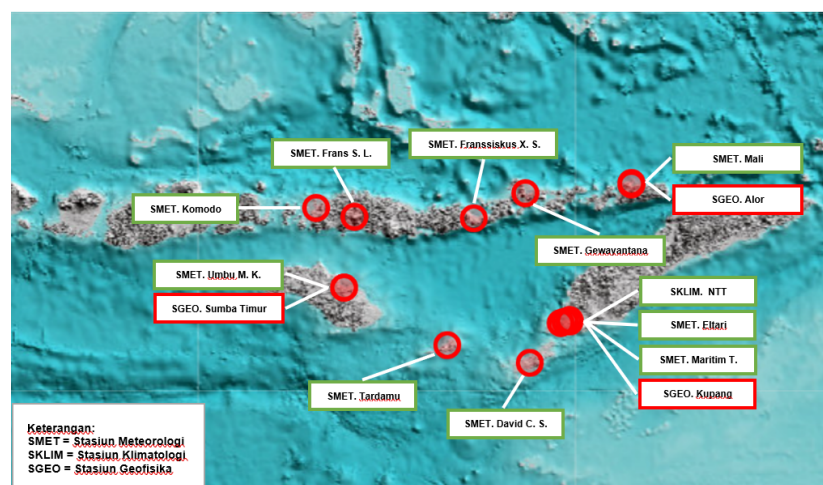
Dalam kerangka penelitian ini, metode Support Vector Regression (SVR) diaplikasikan dengan tujuan memprediksi data temperatur yang hilang di wilayah NTT. SVR merupakan salah satu metode *machine learning* di mana SVR dirancang untuk menyelesaikan masalah dari fungsi regresi untuk menghasilkan Keputusan berdasarkan kombinasi linier dari variabel yang diinputkan (Rahmi & Helma, 2023). Support Vector Regression (SVR) adalah model regresi yang merupakan pengembangan dari SVM, SVR mampu mengatasi masalah *overfitting* dengan cara meminimalkan batas atas dari *generalization error* (Ginting et al., 2021). SVR bertujuan untuk menemukan fungsi $f(x)$ dengan cara memetakan suatu data yang lebih tinggi memakai fungsi kernel. Pemilihan metode SVR dipilih dengan alasan kemampuannya untuk menangani data yang memiliki struktur yang kompleks dan beragam.

2. METODE PENELITIAN

Penelitian ini mengadopsi pendekatan kuantitatif dengan memanfaatkan bahasa pemrograman Python. Salah satu alat yang digunakan dalam proses analisis adalah Google Colaboratory (Google Colab), sebuah platform berbasis *cloud* yang dikembangkan oleh Google Research. Platform ini memungkinkan pengguna untuk menulis, menjalankan, serta berbagi kode Python secara gratis, sehingga sangat mendukung pelaksanaan eksperimen dan analisis data secara efisien (Gat et al., 2023).

2.1 Sumber Data

Data yang digunakan dalam penelitian ini merupakan data sekunder berupa data temperatur maksimum (derajat Celcius) yang bersumber dari data stasiun dan data *Reanalysis* ERA-5. Data stasiun diperoleh dari publikasi online yang dikeluarkan oleh Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) melalui laman <https://dataonline.bmkg.go.id> dengan periode data temperatur harian yaitu 1 Januari 2020 hingga 31 Desember 2023. Wilayah yang menjadi lokasi penelitian adalah Provinsi NTT dengan 22 Kabupaten/Kota dan 14 stasiun pengamatan. Dari data yang diperoleh, tidak semua stasiun memiliki data temperatur maksimum. Oleh karena itu, pada penelitian ini hanya melakukan pengamatan di enam stasiun seperti yang disajikan pada Tabel 1. Sementara itu, data *Reanalysis* ERA-5 tersedia di situs Climate Data Store (CDS) yang dapat di akses melalui laman <https://cds.climate.copernicus.eu/> dalam format NetCDF dengan periode data temperatur yang diambil yaitu dari 1 Januari 2020 hingga 31 Desember 2023.



Gambar 1 Titik Lokasi Pengamatan



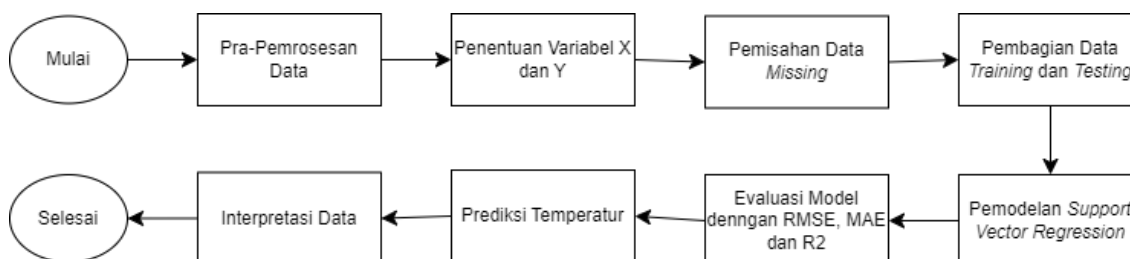
Berdasarkan Gambar 1 menampilkan titik lokasi pengamatan di Wilayah Nusa Tenggara Timur, terdapat 14 titik pengamatan, dengan 3 stasiun berwarna merah yang berarti tidak terdapat pengamatan pada stasiun pengamatan tersebut. Di sini diambil 6 stasiun pengamatan untuk dilakukan penelitian dengan pertimbangan stasiun dengan titik yang berdekatan akan diambil salah satu secara acak, untuk stasiun yang digunakan dapat dilihat pada Tabel 1. Pada Tabel 1 terdapat kolom Y, X1, X2, X3, X4, X5 sebagai struktur data yang digunakan. Variabel y merupakan variabel respon yang berasal dari data temperatur maksimum data stasiun dan variabel X1, X2, X3, X4, X5 merupakan variabel prediktor dari data temperatur maksimum *Reanalysis* ERA-5. Untuk data yang dijadikan sampel merupakan data pada 1 Januari 2020 pada setiap stasiun.

Tabel 1 Nama Stasiun Pengamatan

| No. | Nama Stasiun Pengamatan | Jenis Data | Y | X1 | X2 | X3 | X4 | X5 | |
|-----|--------------------------------------|-----------------------|------------------|------|-------|-------|-------|-------|-------|
| 1 | Meteorologi Fransiskus Xaverius Seda | Latitude Longitude | -8,63 122,23 | 32.8 | 28.31 | 28.14 | 28.59 | 28.64 | 28.90 |
| 2 | Meteorologi Komodo | Latitude Longitude | -8,48 119,88 | 31.6 | 28.70 | 27.68 | 28.17 | 28.87 | 27.26 |
| 3 | Meteorologi Mali | Latitude Longitude | -8,13 124,59 | 32.9 | 28.35 | 27.90 | 27.76 | 26.56 | 29.56 |
| 4 | Meteorologi Tardamu | Latitude Longitude | -10,5 -9,66 | 32 | 28.86 | 29.09 | 28.78 | 29.05 | 29.04 |
| 5 | Meteorologi Umbu Mehang Kunda | Latitude Longitude | -9,66 120,29 | 33.8 | 29.01 | 28.39 | 31.06 | 28.62 | 29.77 |
| 6 | Meteorologi Maritim Tenau | Latitude Longitude | -10.18 123.52 | NaN | 31.23 | 28.81 | 29.53 | 31.42 | 29.32 |

2.2 Langkah Analisis Data

Tahapan penelitian yang dilakukan dalam studi ini digambarkan pada Gambar 2. Penelitian dimulai dengan tahap pra-pemrosesan data, yang mencakup data dari stasiun BMKG dan data *Reanalysis* ERA-5. Untuk data stasiun, proses diawali dengan pengumpulan data dari enam stasiun pengamatan. Sementara itu, pada data *Reanalysis* ERA-5, setelah data dikumpulkan dalam format NetCDF, dilakukan pencarian nilai temperatur maksimum menggunakan algoritma machine learning, kemudian dilakukan konversi skala temperatur dari Kelvin ke Celsius dengan cara mengurangkan 273,15. Selanjutnya, satuan waktu data diubah dari per jam menjadi harian agar selaras dengan data dari stasiun. Setelah semua data memiliki satuan waktu harian, format file diubah dari NetCDF menjadi CSV agar dapat digunakan dalam tahap pemrosesan berikutnya.

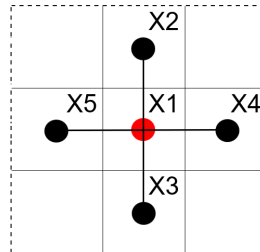


Gambar 2 Tahapan Penelitian

Penentuan variabel Y yaitu sebagai variable respon menggunakan data temperatur maksimum yang didapat dari enam stasiun di Provinsi NTT. Pemilihan enam stasiun ini didasarkan pada



ketersediaan data, karena terdapat beberapa data *missing* di stasiun pengamatan lainnya. Sedangkan untuk variabel X sebagai variabel prediktor menggunakan data temperatur maksimum *Reanalysis* ERA-5 pada *grid* yang berdekatan dengan titik variabel respons. Pada Gambar 3 merupakan hasil penomoran *grid* dengan bagian merah yang berada ditengah adalah *grid* yang tepat pada titik lokasi pengamatan.



Gambar 3 Penentuan *Grid* di Setiap Lokasi Pengamatan

Pada tahap pemisahan data *missing*, data dibagi menjadi dua kelompok, yaitu data yang memiliki nilai lengkap (tersedia) dan data yang mengandung nilai hilang (*missing*). Pemisahan ini bertujuan untuk memastikan bahwa model hanya dilatih menggunakan data yang valid dan lengkap. Data yang tersedia akan digunakan sebagai data latih (*training data*) dalam proses pembangunan model prediksi, sedangkan data yang memiliki nilai hilang akan digunakan untuk evaluasi performa model setelah pelatihan.

Pembagian data *training* dan *testing* menggunakan pembagian 75% untuk data *training* dan 25% data *testing* dari total 1.461 data temperatur maksimum yang tersedia. Alasan pemilihan rasio ini adalah karena beberapa stasiun pengamatan memiliki data yang hilang (*missing*) dalam jumlah yang cukup signifikan. Dengan menggunakan lebih banyak data untuk pelatihan (*training*), model dapat mempelajari pola data dengan lebih baik, sementara tetap menyisakan cukup data untuk pengujian (*testing*) guna mengevaluasi kinerja model secara akurat.

Pemodelan algoritma Support Vector Reggresion (SVR) menggunakan bahasa pemrograman Python karena memiliki *library* dan *framework* yang efisien, mudah digunakan dan *open source*. Pada penggunaan metode SVR biasanya dilakukan penentuan model parameter terbaik, pada penelitian ini penentuan model parameter dilakukan dengan optimasi *Grid Search*, *Grid Search Optimization* akan menguji berbagai kombinasi model dan memilih kombinasi dengan rata-rata kesalahan terkecil (Elsa, 2023; Syahfitri et al., 2024). Beberapa *hyperparameter* yang digunakan untuk *Grid Search* yaitu C, epsilon, dan gama yang diuji dengan nilai [1, 5, 10, 15] dan [0,1, 0,3, 0,6, 0,9] dan fungsi kernel yang digunakan merupakan fungsi kernel Radial Basis Function (RBF) sebagaimana dijabarkan pada Pers. (1) (Ramedani et al., 2014).

$$k(x_i, x_j) = \exp\left(-\frac{d(x_i, x_j)^2}{2l^2}\right) \quad (1)$$

2.2.1 Evaluasi Model

Evaluasi model menggunakan beberapa kriteria seperti Root Mean Square Error (RMSE) dan Mean Absolute Error (MAE) dengan hasil eror yang terkecil dan R2 Score dengan hasil eror yang paling besar. Adapun persamaan yang dapat digunakan untuk menghitung nilai evaluasi model seperti di bawah ini:

1) Root Mean Squared Error (RMSE)

RMSE menghitung akar kuadrat dari rata-rata kuadrat *error*, memberikan bobot yang lebih besar pada *error* yang lebih besar. Nilai RMSE yang lebih kecil menunjukkan bahwa model memiliki prediksi yang lebih akurat. Rumus perhitungan RMSE dijabarkan pada Pers. (2) (Hodson, 2022).



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

2) Mean Absolute Error (MAE)

MAE menghitung rata-rata dari selisih absolut antara nilai yang diprediksi dan nilai aktual dan memberikan gambaran tentang seberapa besar kesalahan prediksi model secara umum. Nilai MAE yang lebih kecil menunjukkan bahwa model memiliki kesalahan prediksi yang lebih kecil dan konsisten. Rumus perhitungan MAE dijabarkan pada Pers. (3) (Nurani et al., 2023).

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{x}_i - x_i| \quad (3)$$

3) Coefficient of Determination (R²)

R² mengukur seberapa baik model dapat menjelaskan variasi dalam data, R² berkisar anantara 0 dan 1. Nilai yang lebih dekat dengan 1 menunjukkan bahwa model dapat menjelaskan sebagian besar variasi data. Rumus perhitungan R² dijabarkan pada Pers. (4) (Baek & Chung, 2023).

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (4)$$

Dalam ketiga persamaan tersebut, n merupakan jumlah total sampel atau data yang digunakan. Simbol \hat{y}_i menunjukkan nilai prediksi dari variabel target y yang dihasilkan oleh model. Selain itu y adalah nilai actual atau sebenarnya dari variabel tersebut.

2.2.2 Imputasi Data

Pada tahap ini, proses dimulai dengan mengidentifikasi data yang hilang atau *missing* untuk memahami skala dan distribusi kekosongan dalam *dataset*, diikuti dengan analisis penyebab kekosongan tersebut guna menentukan apakah data yang hilang terjadi secara acak atau memiliki pola tertentu. Setelah penyebab diidentifikasi, selanjutnya yaitu melakukan implementasi untuk menggantikan data yang hilang dan hasilnya akan dievaluasi untuk memastikan bahwa imputasi yang dilakukan menampilkan hasil secara akurat.

3. HASIL DAN PEMBAHASAN

3.1 Pemodelan dan Prediksi Temperatur Maksimum Stasiun Tardamu

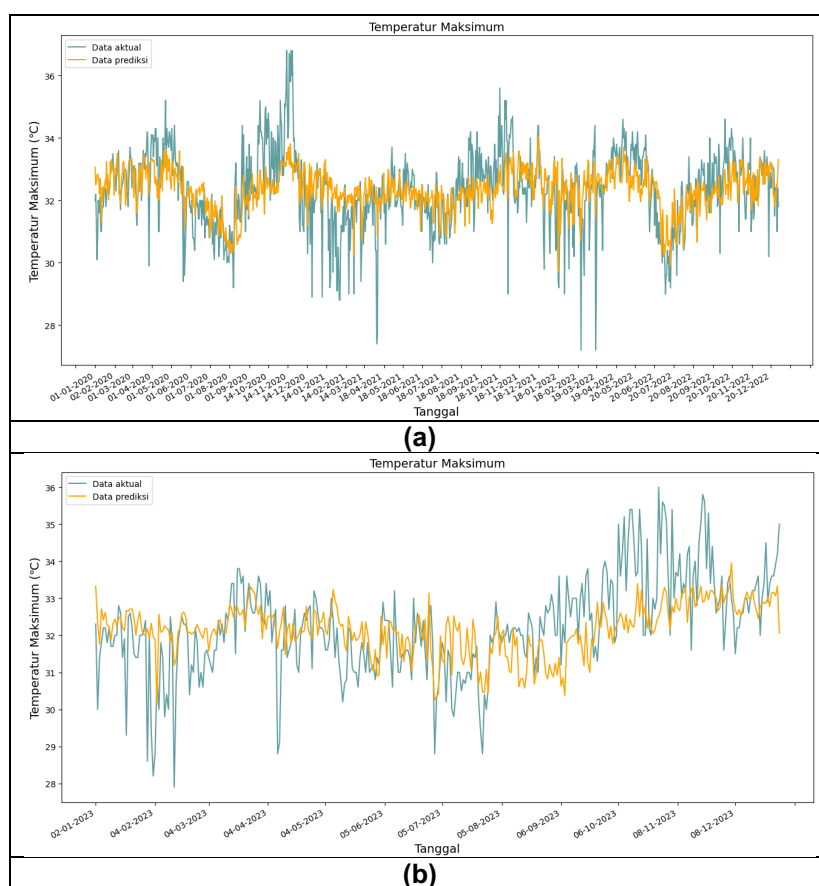
Bagian ini membahas pemodelan dan prediksi temperatur maksimum di Stasiun Tardamu dengan menggunakan berbagai kombinasi parameter untuk memperoleh model yang paling optimal. Pendekatan yang digunakan melibatkan eksplorasi parameter C (C), ε (epsilon), dan γ (gamma) yang berperan dalam mengatur kinerja model. Berdasarkan Tabel 2, hasil evaluasi menunjukkan bahwa kombinasi parameter terbaik diperoleh pada $C = 5$, $\varepsilon = 0,3$, dan $\gamma = 1$, dengan nilai RMSE sebesar 0,994, MAE sebesar 0,721, dan R^2 sebesar 0,356. Meskipun nilai R^2 tergolong rendah, kombinasi ini memberikan hasil terbaik dibandingkan parameter lainnya dalam memodelkan temperatur maksimum di stasiun tersebut.

Untuk memperjelas hasil evaluasi, Gambar 4 menyajikan grafik perbandingan antara data aktual dan hasil prediksi pada data training dan data testing. Pada Gambar 4(a), terlihat bahwa pola prediksi cukup mirip dengan data aktual, meskipun tidak menunjukkan fluktuasi besar yang sebenarnya terjadi dalam data. Sementara itu, Gambar 4(b) memperlihatkan bahwa pada awal tahun 2022 pola data aktual dan prediksi relatif sama, namun mulai Mei 2023 terjadi perbedaan pola yang semakin mencolok, terutama pada bulan September yang menunjukkan deviasi signifikan. Hal ini mengindikasikan bahwa model masih memiliki keterbatasan dalam menangkap dinamika data yang berubah secara musiman atau ekstrem.



Tabel 2 Evaluasi Model Stasiun Tardamu

| Range Parameter | Kriteria Data Training | | | Kriteria Data Testing | | |
|---|------------------------|-------|-------|-----------------------|-------|-------|
| | RMSE | MAE | R2 | RMSE | MAE | R2 |
| $C = 1, 5, 10, 15$ $\varepsilon = 1, 5, 10, 15$ $\gamma = 1, 5, 10, 15$ | 1,000 | 0,754 | 0,347 | 1,121 | 0,841 | 0,275 |
| $C = 1, 5, 10, 15$ $\varepsilon = 0,1, 0,3, 0,6, 0,9$ $\gamma = 0,1, 0,3, 0,6, 0,9$ | 1,029 | 0,755 | 0,310 | 1,127 | 0,842 | 0,267 |
| $C = 1, 5, 10, 15$ $\varepsilon = 0,1, 0,3, 0,6, 0,9$ $\gamma = 1, 5, 10, 15$ | 0,994 | 0,721 | 0,356 | 1,154 | 0,865 | 0,232 |
| Parameter Terbaik | | | | | | |
| $C = 5$ $\varepsilon = 0,3$ $\gamma = 1$ | 0,994 | 0,721 | 0,356 | 1,154 | 0,865 | 0,232 |

Gambar 4 Evaluasi Model Stasiun Tardamu Data *Training* (a) dan Data *Testing* (b)

3.2 Pemodelan dan Prediksi Temperatur Maksimum Stasiun Umu Meheng Kunda

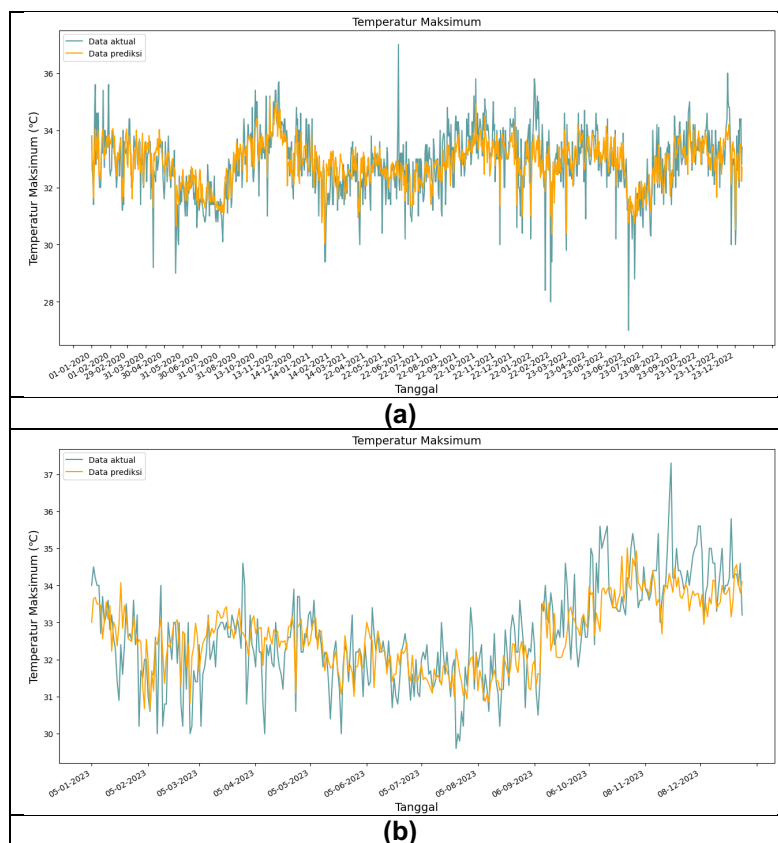
Bagian ini membahas pemodelan dan prediksi temperatur maksimum di Stasiun Umu Meheng Kunda dengan menggunakan berbagai kombinasi parameter melalui pendekatan eksploratif terhadap parameter C (C), ε (epsilon), dan γ (gamma) untuk memperoleh model yang paling optimal berdasarkan kriteria evaluasi tertentu. Berdasarkan Tabel 3, kombinasi parameter terbaik diperoleh pada $C = 1$, $\varepsilon = 0,6$, dan $\gamma = 0,1$, dengan nilai RMSE sebesar 0,718, MAE sebesar



0,551, dan nilai R^2 tertinggi sebesar 0,588. Hasil ini menunjukkan bahwa model memiliki performa yang cukup baik dalam memodelkan temperatur maksimum di stasiun tersebut.

Tabel 3 Evaluasi Model Stasiun Umbu Mehang Kunda

| <i>Range Parameter</i> | <i>Kriteria Data Training</i> | | | <i>Kriteria Data Testing</i> | | |
|---|-------------------------------|-------|-------|------------------------------|-------|-------|
| | RMSE | MAE | R2 | RMSE | MAE | R2 |
| $C = 0,1, 0,3, 0,6, 0,9$ $\varepsilon = 0,1, 0,3, 0,6, 0,9$ $\gamma = 1, 5, 10, 15$ $C = 1, 5, 10, 15$ | 0,723 | 0,554 | 0,582 | 0,868 | 0,656 | 0,558 |
| $\varepsilon = 0,1, 0,3, 0,6, 0,9$ $\gamma = 1, 5, 10, 15$ $C = 1, 5, 10, 15$ | 0,718 | 0,551 | 0,588 | 0,866 | 0,656 | 0,560 |
| $\varepsilon = 1, 5, 10, 15$ $\gamma = 1, 5, 10, 15$ | 0,743 | 0,582 | 0,559 | 0,874 | 0,668 | 0,552 |
| Parameter Terbaik | | | | | | |
| $C = 1$ $\varepsilon = 0,6$ $\gamma = 1$ | 0,718 | 0,551 | 0,588 | 0,866 | 0,656 | 0,560 |



Gambar 5 Evaluasi Model Stasiun Umbu Mehang Kunda Data *Training* (a) dan Data *Testing* (b)

Gambar 5 memperlihatkan hasil prediksi model terhadap data training dan testing. Pada Gambar 5(a), prediksi cukup mendekati data aktual, menandakan kemampuan model mengikuti tren historis dengan baik. Namun, pada Gambar 5(b), terdapat beberapa periode di mana prediksi menyimpang dari data aktual, mengindikasikan keterbatasan model dalam menangkap variasi



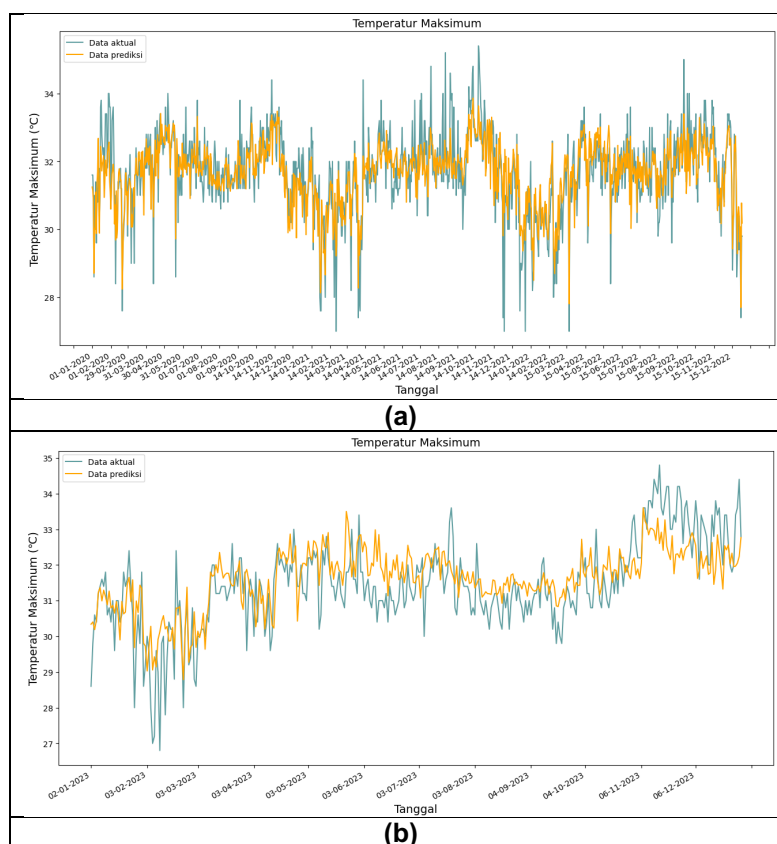
kompleks atau perubahan musiman. Secara keseluruhan, model ini cukup representatif untuk memprediksi temperatur maksimum di lokasi tersebut.

3.3 Pemodelan dan Prediksi Temperatur Maksimum Stasiun Komodo

Bagian ini membahas pemodelan dan prediksi temperatur maksimum di Stasiun Komodo dengan mengeksplorasi kombinasi parameter C (C), ε (epsilon), dan γ (gamma) untuk mendapatkan model optimal. Berdasarkan Tabel 4, kombinasi terbaik adalah $C = 1$, $\varepsilon = 0,3$, dan $\gamma = 1$, dengan RMSE 0,691, MAE 0,520, dan R^2 0,681, menunjukkan performa prediksi yang baik.

Tabel 4 Evaluasi Model Stasiun Komodo

| Range Parameter | Kriteria Data Training | | | Kriteria Data Testing | | |
|--|------------------------|-------|-------|-----------------------|-------|-------|
| | RMSE | MAE | R2 | RMSE | MAE | R2 |
| C = 0,1, 0,3, 0,6, 0,9 ε = 0,1, 0,3, 0,6, 0,9 γ = 1, 5, 10, 15 C = 1, 5, 10, 15 | 0,696 | 0,523 | 0,676 | 0,862 | 0,674 | 0,523 |
| ε = 0,1, 0,3, 0,6, 0,9 γ = 0,1, 0,3, 0,6, 0,9 C = 1, 5, 10, 15 | | | | | | |
| ε = 0,1, 0,3, 0,6, 0,9 γ = 0,1, 0,3, 0,6, 0,9 C = 1, 5, 10, 15 | | | | | | |
| ε = 0,1, 0,3, 0,6, 0,9 γ = 1, 5, 10, 15 | | | | | | |
| C = 0,1, 0,3, 0,6, 0,9 ε = 0,1, 0,3, 0,6, 0,9 γ = 1, 5, 10, 15 C = 1, 5, 10, 15 | 0,691 | 0,520 | 0,681 | 0,861 | 0,674 | 0,524 |
| ε = 0,1, 0,3, 0,6, 0,9 γ = 0,1, 0,3, 0,6, 0,9 C = 1, 5, 10, 15 | | | | | | |
| ε = 0,1, 0,3, 0,6, 0,9 γ = 0,1, 0,3, 0,6, 0,9 C = 1, 5, 10, 15 | | | | | | |
| ε = 0,1, 0,3, 0,6, 0,9 γ = 1, 5, 10, 15 | | | | | | |
| Parameter Terbaik | | | | | | |
| C = 1 ε = 0,3 γ = 1 | 0,691 | 0,520 | 0,681 | 0,861 | 0,674 | 0,524 |



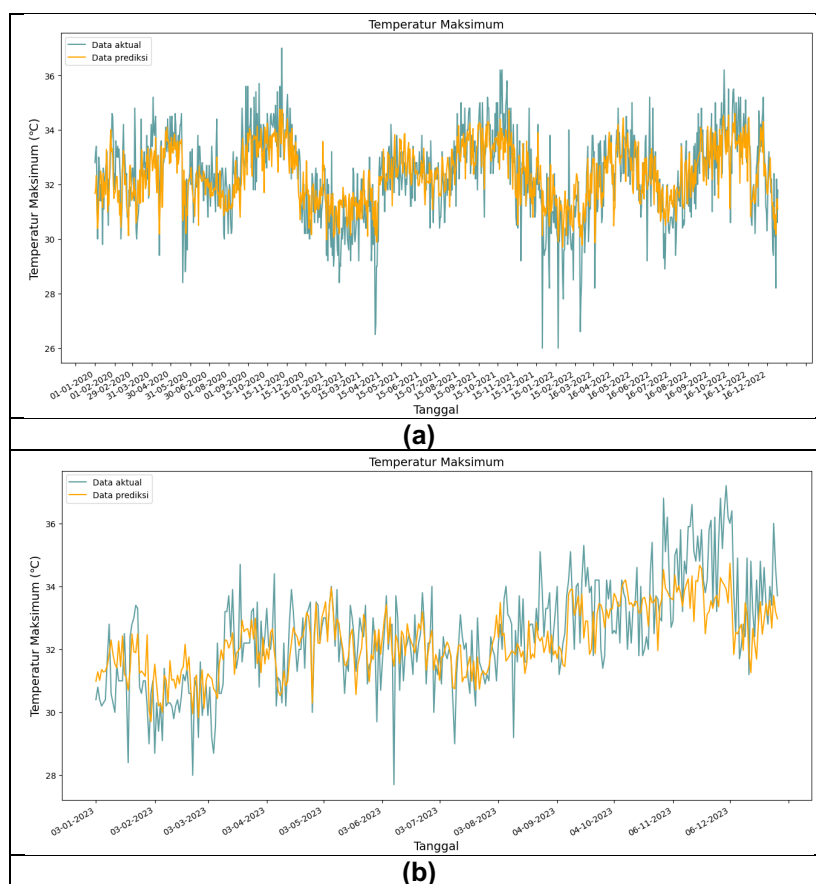
Gambar 6 Evaluasi Model Stasiun Komodo Data *Training* (a) dan Data *Testing* (b)



Gambar 6 memperlihatkan hasil prediksi pada data training dan testing. Pada Gambar 6(a), prediksi cukup konsisten dengan data aktual. Namun, pada Gambar 6(b), prediksi sedikit lebih tinggi dari data aktual di Februari 2023, sementara mulai November 2023 data aktual melebihi prediksi. Hal ini menunjukkan keterbatasan model dalam menangkap dinamika musiman dan perubahan pola mendadak.

3.4 Pemodelan dan Prediksi Temperatur Maksimum Stasiun Fransiskus Xavierius Seda

Bagian ini membahas proses pemodelan dan prediksi temperatur maksimum di Stasiun Fransiskus Xavierius Seda dengan pendekatan eksploratif terhadap berbagai kombinasi parameter. Gambar 7 memperlihatkan visualisasi hasil prediksi model dibandingkan dengan data aktual pada data training dan testing. Pada gambar tersebut, model menunjukkan kemampuan generalisasi yang cukup baik, dengan prediksi yang mendekati data aktual terutama pada data training, meskipun pada data testing terdapat beberapa deviasi akibat variasi ekstrem atau perubahan musiman.



Gambar 7 Evaluasi Model Stasiun Fransiskus Xavierius Seda Data *Training* (a) dan Data *Testing* (b)

Selanjutnya, berdasarkan Tabel 5, kombinasi parameter optimal diperoleh pada $C = 1$, $\varepsilon = 0,9$, dan $\gamma = 0,1$, dengan nilai RMSE sebesar 0,966, MAE sebesar 0,765, dan R^2 sebesar 0,601. Nilai-nilai ini mengindikasikan bahwa model berhasil mempelajari hubungan antara fitur masukan dan temperatur maksimum dengan cukup baik. Hasil visualisasi dan evaluasi kuantitatif ini bersama-sama memberikan gambaran yang komprehensif mengenai performa model pada stasiun tersebut.

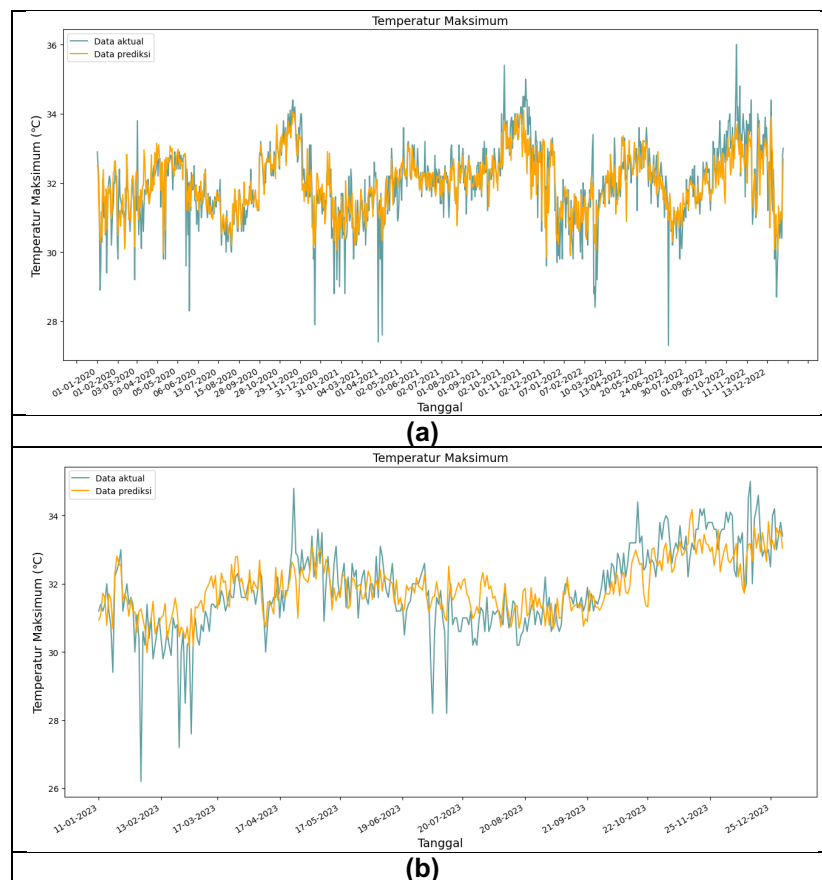


Tabel 5 Evaluasi Model Stasiun Fransiskus Xaverius Seda

| Range Parameter | Kriteria Data Training | | | Kriteria Data Testing | | |
|---|------------------------|-------|-------|-----------------------|-------|-------|
| | RMSE | MAE | R2 | RMSE | MAE | R2 |
| $C = 0,1, 0,3, 0,6, 0,9$ $\varepsilon = 0,1, 0,3, 0,6, 0,9$ $\gamma = 1, 5, 10, 15$ $C = 1, 5, 10, 15$ | 0,971 | 0,767 | 0,597 | 1,175 | 0,928 | 0,508 |
| $\varepsilon = 0,1, 0,3, 0,6, 0,9$ $\gamma = 1, 5, 10, 15$ $C = 1, 5, 10, 15$ | 0,966 | 0,765 | 0,601 | 1,174 | 0,927 | 0,509 |
| $\varepsilon = 1, 5, 10, 15$ $\gamma = 1, 5, 10, 15$ | 0,969 | 0,768 | 0,599 | 1,175 | 0,928 | 0,508 |
| Parameter Terbaik | | | | | | |
| $C = 1$ $\varepsilon = 0,9$ $\gamma = 1$ | 0,966 | 0,765 | 0,601 | 1,174 | 0,927 | 0,509 |

3.5 Pemodelan dan Prediksi Temperatur Maksimum Stasiun Mali

Bagian ini membahas pemodelan temperatur maksimum di Stasiun Mali dengan mengeksplorasi berbagai kombinasi parameter. Gambar 8 menunjukkan perbandingan antara data aktual dan prediksi pada data training dan testing. Gambar 8(a) memperlihatkan bahwa model cukup baik mengikuti pola data, meskipun kurang menangkap fluktuasi tajam. Sementara itu, Gambar 8(b) menunjukkan model belum mampu memprediksi nilai ekstrem, seperti temperatur di bawah 30°C, meskipun data aktual mencapai 26°C.



Gambar 8 Evaluasi Model Stasiun Mali Data Training (a) dan Data Testing (b)



Tabel 6 menyajikan hasil evaluasi kombinasi parameter, dengan konfigurasi terbaik pada $C = 1$, $\varepsilon = 0,1$, dan $\gamma = 1$. Model ini menghasilkan RMSE 0,593, MAE 0,390, dan R^2 0,704. Hasil ini menunjukkan performa model cukup akurat dan lebih baik dibandingkan stasiun lainnya, meski masih perlu peningkatan dalam menangani variasi ekstrem.

Tabel 6 Evaluasi Model Stasiun Mali

| <i>Range Parameter</i> | <i>Kriteria Data Training</i> | | | <i>Kriteria Data Testing</i> | | |
|---|-------------------------------|-------|-------|------------------------------|-------|-------|
| | RMSE | MAE | R2 | RMSE | MAE | R2 |
| $C = 0,1, 0,3, 0,6, 0,9$ $\varepsilon = 0,1, 0,3, 0,6, 0,9$ $\gamma = 1, 5, 10, 15$ | 0,599 | 0,395 | 0,698 | 0,835 | 0,619 | 0,533 |
| $C = 1, 5, 10, 15$ $\varepsilon = 0,1, 0,3, 0,6, 0,9$ $\gamma = 1, 5, 10, 15$ | 0,593 | 0,390 | 0,704 | 0,834 | 0,619 | 0,535 |
| $C = 1, 5, 10, 15$ $\varepsilon = 1, 5, 10, 15$ $\gamma = 0,1, 0,3, 0,6, 0,9$ | 0,663 | 0,512 | 0,631 | 0,729 | 0,541 | 0,644 |
| Parameter Terbaik | | | | | | |
| $C = 1$ $\varepsilon = 0,1$ $\gamma = 1$ | 0,593 | 0,390 | 0,704 | 0,834 | 0,619 | 0,535 |

3.6 Pemodelan dan Prediksi Temperatur Maksimum Stasiun Maritim Tenau

Bagian ini membahas pemodelan dan prediksi temperatur maksimum di Stasiun Maritim Tenau dengan mengeksplorasi berbagai kombinasi parameter untuk memperoleh model yang paling optimal. Pendekatan yang digunakan melibatkan pengujian parameter seperti C (C), ε (epsilon), dan γ (gamma) berdasarkan beberapa metrik evaluasi. Berdasarkan hasil evaluasi yang ditampilkan pada Tabel 7, kombinasi parameter terbaik diperoleh pada $C = 5$, $\varepsilon = 0,9$, dan $\gamma = 1$, yang menghasilkan nilai error terkecil. Dengan konfigurasi ini, model menghasilkan nilai RMSE sebesar 0,685, MAE sebesar 0,585, dan nilai R^2 tertinggi sebesar 0,721, yang menunjukkan bahwa model memiliki performa prediktif yang cukup baik untuk stasiun ini.

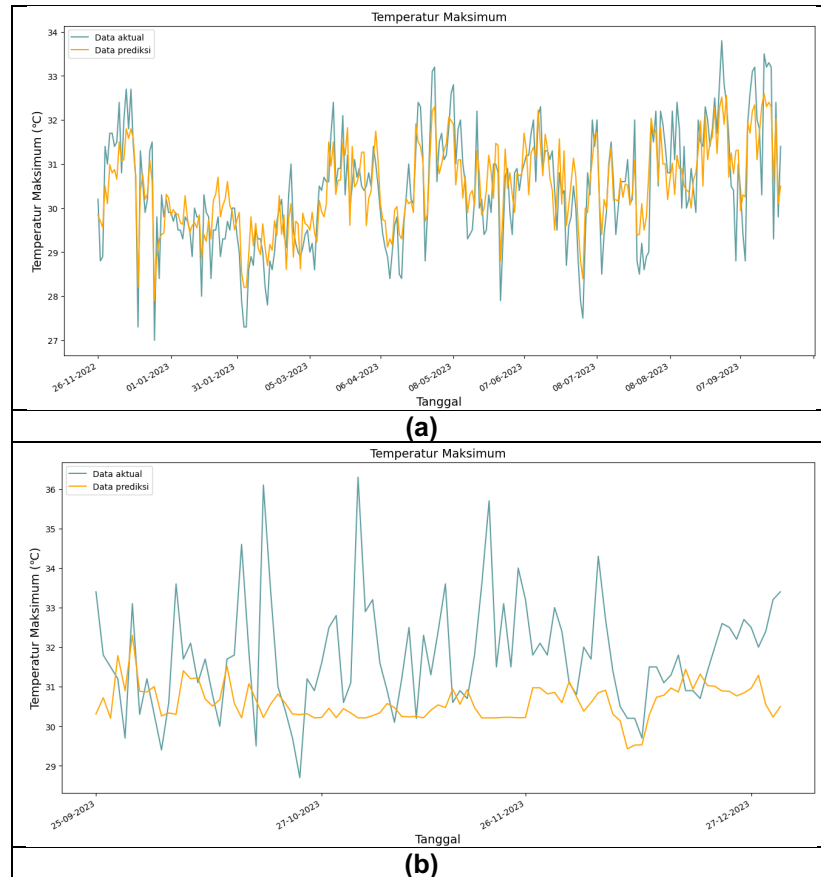
Tabel 7 Evaluasi Model Stasiun Maritim Tenau

| <i>Range Parameter</i> | <i>Kriteria Data Training</i> | | | <i>Kriteria Data Testing</i> | | |
|---|-------------------------------|-------|-------|------------------------------|-------|--------|
| | RMSE | MAE | R2 | RMSE | MAE | R2 |
| $C = 0,1, 0,3, 0,6, 0,9$ $\varepsilon = 0,1, 0,3, 0,6, 0,9$ $\gamma = 1, 5, 10, 15$ | 0,714 | 0,568 | 0,698 | 1,862 | 1,381 | -0,797 |
| $C = 1, 5, 10, 15$ $\varepsilon = 0,1, 0,3, 0,6, 0,9$ $\gamma = 1, 5, 10, 15$ | 0,685 | 0,585 | 0,721 | 1,897 | 1,409 | -0,866 |
| $C = 1, 5, 10, 15$ $\varepsilon = 1, 5, 10, 15$ $\gamma = 1, 5, 10, 15$ | 0,720 | 0,611 | 0,693 | 1,905 | 1,419 | -0,882 |
| Parameter Terbaik | | | | | | |
| $C = 1$ $\varepsilon = 0,9$ $\gamma = 1$ | 0,685 | 0,585 | 0,721 | 1,897 | 1,409 | -0,866 |

Untuk memperjelas hasil evaluasi, Gambar 9 menyajikan visualisasi perbandingan antara data aktual dan hasil prediksi pada data training dan testing. Pada Gambar 9(a), tampak bahwa prediksi model cukup mengikuti pola data aktual pada data training. Namun, pada Gambar 9(b), terlihat bahwa data aktual mengalami fluktuasi naik-turun yang tidak beraturan, sedangkan hasil



prediksi cenderung linier dan kurang mampu menangkap dinamika ekstrem dari temperatur maksimum. Hal ini menunjukkan bahwa meskipun performa model secara umum cukup baik, masih terdapat keterbatasan dalam menangkap variasi ekstrem yang muncul pada data aktual.



Gambar 9 Evaluasi Model Stasiun Maritim Tenau Data *Training* (a) dan Data *Testing* (b)

4. KESIMPULAN

Hasil penelitian menunjukkan bahwa metode Support Vector Regression (SVR) efektif dalam memprediksi temperatur maksimum harian di wilayah Nusa Tenggara Timur. Metode ini mampu menangani pola data yang terbatas (tidak lengkap) dan bersifat non-linear dengan baik. Evaluasi menggunakan metrik RMSE, MAE, dan R^2 menunjukkan bahwa SVR menghasilkan prediksi yang mendekati nilai sebenarnya dengan tingkat kesalahan yang rendah, terutama saat menggunakan kombinasi parameter terbaik. Parameter terbaik untuk setiap stasiun berada dalam rentang $C = [1, 5, 10, 15]$, $\varepsilon = [0,1, 0,3, 0,6, 0,9]$, dan $\gamma = [1, 5, 10, 15]$.

Meskipun SVR terbukti efektif dalam mengatasi imputasi data missing temperatur maksimum di wilayah NTT, masih terdapat peluang untuk pengembangan lebih lanjut. Penelitian selanjutnya disarankan membandingkan metode SVR dengan teknik regresi lain guna menentukan metode paling optimal dalam menangani data klimatologis yang tidak lengkap. Selain itu, eksplorasi parameter lain, seperti jenis kernel atau teknik optimasi yang berbeda, dapat dipertimbangkan untuk meningkatkan akurasi prediksi dan kemampuan generalisasi model.

DAFTAR PUSTAKA

Agwil, W., Agustina, D., Fransiska, H., & Hasani, I. A. (2024). Meningkatkan Kinerja Model Klasifikasi Curah Hujan Melalui Penanggulangan Missing Value dengan Imputasi Berbasis



Artikel ini didistribusikan mengikuti lisensi Atribusi-NonKomersial CC BY-NC sebagaimana tercantum pada <https://creativecommons.org/licenses/by-nc/4.0/>.

- Model. *INNOVATIVE: Journal of Social Science Research*, 4(1), 11773–11783. <https://doi.org/10.31004/innovative.v4i1.9158>
- Baek, J. W., & Chung, K. (2023). Multi-Context Mining-Based Graph Neural Network for Predicting Emerging Health Risks. *IEEE Access*, 11, 15153–15163. <https://doi.org/10.1109/ACCESS.2023.3243722>
- BPK RI Perwakilan Provinsi Nusa Tenggara Timur. (2024). *Pemerintah Provinsi NTT*. BPK RI. <https://ntt.bpk.go.id/pemerintah-provinsi-ntt/>
- Elsa, R. (2023). *Penerapan Metode Support Vector Regression (SVR) Menggunakan Kernel Linear, Polinomial, dan Radial dengan Grid Search Optimization* [Universitas Lampung]. <https://digilib.unila.ac.id/74539/3/SKRIPSI%20TANPA%20BAB%20PEMBAHASAN.pdf>
- Gat, G., Hidayatullah, A., & Berliana, A. (2023). Workshop Pengenalan Dasar Pemrograman Python dengan Google Colaboratory. *Prosiding ABDIMAS CORISINDO, 2023*, 65–70. <https://ojs.stmikpontianak.ac.id/index.php/pengabdian/article/view/74>
- Ginting, L. M., Sigiuro, M. MT., Manurung, E. D., & Sinurat, J. J. P. (2021). Perbandingan Metode Algoritma Support Vector Regression dan Multiple Linear Regression untuk Memprediksi Stok Obat. *Journal of Applied Technology and Informatics Indonesia*, 1(2), 29–34. <https://doi.org/10.54074/jati.v1i2.36>
- Harvian, K. A., & Yuhan, R. J. (2021). Kajian Perubahan Iklim Terhadap Ketahanan Pangan. *Seminar Nasional Official Statistics*, 1052–1061. <https://doi.org/10.34123/semnasoffstat.v2020i1.593>
- Hodson, T. O. (2022). Root-Mean-Square Error (RMSE) or Mean Absolute Error (MAE): When to Use Them or Not. *Geoscientific Model Development*, 15(14), 5481–5487. <https://doi.org/10.5194/gmd-15-5481-2022>
- Jabir, S. R., Azis, H., Widyawati, D., & Tenripada, A. U. (2023). Prediksi Potensi Donatur Menggunakan Model Logistic Regression. *Indonesian Journal of Data and Science*, 4(1), 31–37. <https://doi.org/10.56705/ijodas.v4i1.64>
- Laia, M. (2023). Analisis Kinerja Algoritma K-Nearest Neighbor Imputation (KNNI) untuk Missing Value pada Klasifikasi Data Mining. *Journal of Informatics, Electrical and Electronics Engineering*, 2(3), 92–98. <https://doi.org/10.47065/jieeee.v2i3.891>
- Malihah, L. (2022). Tantangan Dalam Upaya Mengatasi Dampak Perubahan Iklim dan Mendukung Pembangunan Ekonomi Berkelanjutan: Sebuah Tinjauan. *Jurnal Kebijakan Pembangunan*, 17(2), 219–232. <https://doi.org/10.47441/jkp.v17i2.272>
- Natalia, F., & Firdaus, I. (2021, August 11). *Kebakaran Lahan di NTT Capai 165 Titik Panas, Stasiun Meteorologi: Ada Praktik Pembakaran Ladang*. Kompas.TV. <https://www.kompas.tv/bisnis/200652/kebakaran-lahan-di-ntt-capai-165-titik-panas-stasiun-meteorologi-ada-praktik-pembakaran-ladang>
- Nurani, A. T., Setiawan, A., & Susanto, B. (2023). Perbandingan Kinerja Regresi Decision Tree dan Regresi Linear Berganda untuk Prediksi BMI pada Dataset Asthma. *Jurnal Sains dan Edukasi Sains*, 6(1), 34–43. <https://doi.org/10.24246/juses.v6i1p34-43>
- Prasetya, M. R. A., Priyatno, A. M., & Nurhaeni. (2023). Penanganan Imputasi Missing Values pada Data Time Series dengan Menggunakan Metode Data Mining. *Jurnal Informasi dan Teknologi*, 5(2), 52–62. <https://doi.org/10.37034/jidt.v5i2.324>
- Pratikno, A., & Rahardjo, M. (2021). Analisis Strategi Peningkatan Kinerja Perusahaan dengan Pendekatan SWOT-Balanced Scorecard pada PT IPSI Karya Abadi. *Jurnal Manajemen Bisnis dan Kewirausahaan*, 5(2), 184–189. <https://doi.org/10.24912/jmbk.v5i2.11229>
- Rahmi, A., & Helma, H. (2023). Portofolio Optimal dengan Mempertimbangkan Prediksi Return Menggunakan Metode Support Vector Regression (SVR) Program Studi Matematika. *Jurnal Pendidikan Tambusai*, 7(3), 23745–23753. <https://doi.org/10.31004/jptam.v7i3.10380>
- Ramedani, Z., Omid, M., Keyhani, A., Shamshirband, S., & Khoshnevisan, B. (2014). Potential of Radial Basis Function Based Support Vector Regression for Global Solar Radiation Prediction. *Renewable and Sustainable Energy Reviews*, 39, 1005–1011. <https://doi.org/10.1016/j.rser.2014.07.108>
- Sefidian, A. M., & Daneshpour, N. (2019). Missing Value Imputation Using a Novel Grey Based Fuzzy C-Means, Mutual Information Based Feature Selection, and Regression Model. *Expert Systems with Applications*, 115, 68–94. <https://doi.org/10.1016/j.eswa.2018.07.057>



- Septiansari, A. P. D., Zakaria, A., Khotimah, S. N., & Romdania, Y. (2021). Analisis Data Curah Hujan yang Hilang dengan Menggunakan Metode Normal Ratio, Inversed Square Distance, Rata-Rata Aljabar, dan Linear Regression (Studi Kasus Data Curah Hujan Beberapa Stasiun Hujan Wilayah Lampung Tengah). *Jurnal Rekayasa Sipil dan Desain*, 9(4), 853–862. <https://doi.org/10.23960/jrsdd.v9i4.2219>
- Syahfitri, N., Amalita, N., Vionanda, D., & Martha, Z. (2024). Forecasting Gold Prices in Indonesia Using Support Vector Regression with the Grid Search Algorithm. *UNP Journal of Statistics and Data Science*, 2(1), 32–39. <https://doi.org/10.24036/ujsds/vol2-iss1/145>

