

Prediksi Curah Hujan di Kabupaten Berau Menggunakan *Support Vector Regression*

Rainfall Prediction in Berau Regency Using Support Vector Regression

Monalisa Randang Patiallo¹, M. Fathurahman^{2,a}, Surya Prangga³, Erlyne Nadhilah Widyaningrum⁴

^{1,2,3,4}Program Studi S-1 Statistika, Jurusan Matematika FMIPA Universitas Mulawarman Samarinda, Indonesia

^{a)} Corresponding author: fathur@fmipa.unmul.ac.id

ABSTRACT

Machine learning is an analytical approach that is able to predict the output of a system based on patterns that have been formed from previous data. One of the machine learning methods used in this research is Support Vector Regression (SVR). SVR is the application of the support vector machine method in the case of regression. The concept of the SVR algorithm is to obtain a function with the minimum error rate so as to produce a good predictive value. The advantage of SVR lies in its ability to handle nonlinear data using the kernel functions. This study aims to determine the results of rainfall prediction in Berau Regency using the SVR method. The data used is rainfall data in Berau Regency from January 2014 to December 2023 as much as 120 data, and uses five predictor variables namely temperature, humidity, air pressure, wind speed, and solar irradiation. The proportion of training and testing data is 80%:20%. The kernel function used is a polynomial kernel with parameter values $C = 0,5; 1; 2; 4$ and $d = 1; 2$. The results showed that the best SVR model to predict rainfall in Berau Regency is the SVR model with parameter values $C = 1$ and $d = 1$. This model provides good prediction performance, with an RMSE value of 0,1786.

Keywords: Rainfall, Machine learning, Prediction, SVR.

1. Pendahuluan

Kemajuan ilmu pengetahuan dan teknologi yang sangat pesat mendorong munculnya berbagai inovasi besar, terutama dalam bidang kecerdasan buatan atau *Artificial Intelligence* (AI). AI mengembangkan beberapa metode untuk menjalankan algoritmanya, salah satunya adalah *machine learning*. *Machine learning* digunakan sebagai alat statistika (*statistical tools*) yang digunakan dalam analisis data. Cara *machine learning* untuk membuat keputusan berdasarkan data yang ada, adalah melalui suatu pendekatan. Pendekatan tersebut terbagi menjadi tiga macam konsep pembelajaran, yaitu *Supervised Learning*, *Unsupervised Learning*, dan *Reinforcement Learning* (Nurhalizah et al., 2024).

Metode *supervised learning* didasarkan pada kumpulan sampel data yang memiliki label dan dikelompokkan lebih lanjut dalam masalah klasifikasi dan regresi. Salah satu metode untuk masalah klasifikasi adalah *Support Vector Machine* (SVM). SVM termasuk dalam metode *supervised learning* yang dikembangkan untuk menyelesaikan masalah klasifikasi dengan mencari *hyperplane* (garis pemisah) terbaik antar kelas. Metode SVM tidak hanya digunakan untuk klasifikasi, tetapi juga dapat dimodifikasi untuk menyelesaikan masalah regresi, yaitu dalam memprediksi nilai numerik. Modifikasi ini dikembangkan dengan metode *Support Vector Regression* (SVR) (Hermawan et al., 2022).

Metode SVR telah banyak digunakan di berbagai bidang untuk menyelesaikan permasalahan dari kasus prediksi dengan tingkat *error* cukup rendah (Siregar, 2022). Salah satu bidang yang erat dengan kegiatan memprediksi adalah bidang klimatologi. Bidang klimatologi adalah cabang ilmu yang mempelajari iklim, termasuk pola cuaca jangka panjang dan perubahan yang terjadi di dalamnya. Salah satu variabel penentu kondisi iklim adalah curah hujan. Curah hujan merupakan jumlah air hujan yang turun pada suatu daerah dalam kurun waktu tertentu (BPS, 2023). Proses terjadinya hujan berhubungan dengan unsur-unsur cuaca lainnya seperti suhu, kelembaban, tekanan udara, kecepatan angin, dan lama penyinaran matahari (Sunarmi et al., 2022). Curah hujan sendiri memiliki keragaman yang besar baik dalam skala ruang maupun skala waktu serta pola data dari curah hujan seringkali menghasilkan data yang *nonlinear*. Metode SVR mampu mengatasi data curah hujan yang *nonlinear*, serta memiliki kinerja yang baik meskipun dengan data yang terbatas (Abdullah & Said, 2025).

Curah hujan dengan intensitas tinggi berpotensi menimbulkan banjir di berbagai wilayah, termasuk Kabupaten Berau, Provinsi Kalimantan Timur, yang memiliki kerentanan tinggi terhadap peristiwa tersebut. Berdasarkan data yang dihimpun oleh Badan Penanggulangan Bencana Daerah (BPBD) Kabupaten Berau, banjir terjadi di beberapa kecamatan saat musim penghujan, seperti pada tahun 2020 terjadi di Kecamatan Segah, Teluk Bayur, dan Tanjung Redeb (Karondia et al., 2022). Selebihnya dalam publikasi BPBD Berau Tahun 2021 dinyatakan bahwa terdapat sembilan bencana-bencana yang akan berpotensi terjadi di Kabupaten Berau Tahun 2020-2024, dimana banjir berada pada urutan pertama, disusul banjir bandang, serta cuaca

ekstrim berada pada urutan ketiga (BPBD Berau, 2021). Penyebab bencana tersebut tidak lain dapat diketahui dari adanya dampak intensitas curah hujan. Oleh karena itu, penting dilakukan penelitian untuk memprediksi curah hujan dengan data meteorologi yang telah dikumpulkan, guna memberikan informasi mengenai curah hujan khususnya pemerintah serta masyarakat setempat di Kabupaten Berau.

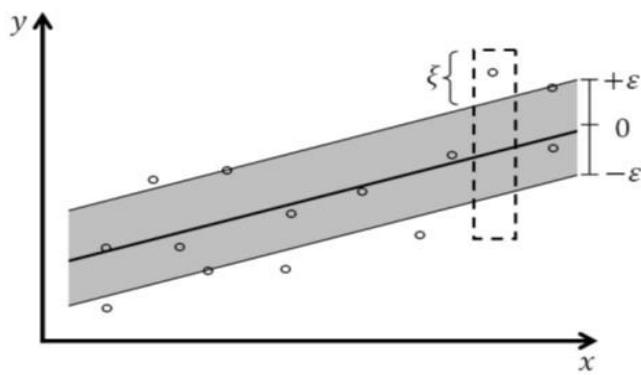
Beberapa penelitian terdahulu yang berkaitan dengan penggunaan metode SVR dalam hal prediksi di antaranya dilakukan oleh Abdullah & Said (2024) yang memprediksi curah hujan harian dan bulanan di India Utara, Aligarh menggunakan kernel linear, radial basis function (RBF), dan polynomial dengan variabel prediktor seperti suhu, titik embun, kelembaban, kecepatan angin, dan tutupan awan. Hasil penelitian tersebut menunjukkan bahwa SVR dengan kernel polynomial memberikan akurasi terbaik untuk prediksi curah hujan bulanan. Terdapat penelitian yang menerapkan SVR dalam prediksi harga saham menggunakan kernel linear, RBF, dan polynomial, dan menemukan bahwa kernel polynomial memberikan akurasi paling baik dengan nilai RMSE yang lebih rendah (Artini et al., 2024). Selanjutnya, terdapat penelitian memprediksi jumlah kasus terkonfirmasi COVID-19 di Indonesia dengan berbagai jenis kernel dan memperoleh hasil bahwa kernel polynomial memberikan nilai R^2 tertinggi (Safira et al., 2023).

Berdasarkan beberapa penelitian tersebut, terlihat bahwa metode SVR dengan variasi kernel mampu memberikan performa yang baik pada berbagai bidang prediksi. Namun, belum banyak penelitian yang secara spesifik menerapkan metode SVR untuk prediksi curah hujan di wilayah Kabupaten Berau, yang memiliki karakteristik iklim dan pola curah hujan berbeda dibandingkan daerah lain di Indonesia maupun Asia Selatan. Oleh karena itu, penelitian ini dilakukan untuk mengisi kesenjangan tersebut dengan mengimplementasikan metode Support Vector Regression (SVR) dalam memodelkan dan memprediksi curah hujan di Kabupaten Berau, serta mengevaluasi performa berbagai jenis kernel dalam konteks data iklim lokal, sehingga dapat memberikan kontribusi empiris terhadap pengembangan model prediksi curah hujan berbasis pembelajaran mesin di Indonesia.

2. Tinjauan Pustaka

2.1 Support Vector Regression

Support Vector Regression (SVR) merupakan pengembangan dari metode *Support Vector Machine* (SVM) yang digunakan pada permasalahan regresi. SVR bertujuan untuk menemukan sebuah fungsi sebagai garis pemisah (*hyperlane*) terbaik yang berupa fungsi regresi sesuai dengan semua input data dengan *error* sekecil mungkin. SVM berfungsi untuk klasifikasi ke dalam dua zona, lain halnya dengan SVR yang justru mengumpulkan data menjadi satu dengan meminimalisasi nilai *epsilon* (ϵ), dengan kata lain SVM membagi menjadi dua, sedangkan SVR memasukkan semuanya menjadi satu (Suyono et al., 2022). Berikut konsep SVR dapat dilihat pada Gambar 1.



Gambar 1. Ilustrasi konsep SVR

Pada Gambar 1, garis diagonal di tengah yang disebut *hyperplane* sedangkan dua garis yang mengapitnya disebut *soft margin*. Jarak antara *hyperplane* dan *soft margin* adalah sebesar ϵ dan titik-titik yang berada di luar dan tepat berada pada batas $+\epsilon$ dan $-\epsilon$ disebut *support vector*, dan untuk titik yang melewati *soft margin* dibutuhkan adanya variabel slack (ξ).

Ide dasar SVR adalah menetukan set data yang dibagi menjadi data *training* dan data *testing*. Misalkan terdapat n set data *training*, (\mathbf{x}_i, y_i) dengan $i = 1, 2, \dots, n$, $\mathbf{x}_i = \{x_1, x_2, \dots, x_n\} \in R^n$ merupakan vektor dalam input *space* dan $y_i = \{y_1, y_2, \dots, y_n\} \in R$ merupakan nilai *output* berdasarkan \mathbf{x}_i yang bersesuaian, maka pada metode SVR persamaan fungsi regresi secara umum dapat dituliskan sebagai berikut (Ishlah et al., 2023):

$$f(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + b \quad (1)$$

dengan $f(\mathbf{x})$ merupakan fungsi regresi yang menunjukkan nilai prediksi, \mathbf{w} merupakan vektor bobot, \mathbf{x} adalah vektor input dan b adalah bias.

Parameter \mathbf{w} dan b adalah koefisien yang harus diprediksi. Agar mendapatkan generalisasi yang baik pada fungsi $f(\mathbf{x})$ maka dapat dilakukan dengan meminimalkan $\|\mathbf{w}\|$ dengan penyelesaian masalah optimasi *hyperlane* dalam bentuk *Quadratic Programming* sebagai berikut (Purwoko et al., 2023):

$$\min \frac{1}{2} \|\mathbf{w}\|^2 \quad (2)$$

dengan syarat $\begin{cases} y_i - \mathbf{w}^T \mathbf{x}_i - b \leq \varepsilon \\ \mathbf{w}^T \mathbf{x}_i + b - y_i \leq \varepsilon \end{cases}$ untuk $i = 1, 2, \dots, n$

dimana nilai $\|\mathbf{w}\|^2$ disebut dengan *regularisasi* yang merupakan fungsi yang diminimumkan agar membuat fungsi setipis (*flat*) mungkin atau suatu teknik yang dapat menghasilkan model lebih sederhana, y_i merupakan nilai aktual periode ke- i , $\mathbf{w}^T \mathbf{x}_i$ adalah hasil perkalian skalar (*dot product*) antara vektor bobot \mathbf{w} dan vektor input untuk periode ke- i , dan ε merupakan skalar atau konstanta yang menyatakan tingkat toleransi.

Pada Persamaan (2) semua fungsi regresi $f(\mathbf{x})$ diasumsikan dapat membaca semua titik (\mathbf{x}_i, y_i) dengan batas toleransi *error* sebesar ε . Setiap data yang lebih besar dari ε akan dikenakan pinalti yang diukur berdasarkan nilai variabel ξ dan dikalikan dengan konstanta C . Hal ini disebut dengan konsep ε -insensitive loss function. Oleh karena itu, masalah optimasi pada persamaan (2) ditransformasikan ke dalam bentuk persamaan (3) (Purwoko et al., 2023).

$$\min \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \quad (3)$$

yang memenuhi: $\begin{cases} y_i - \mathbf{w}^T \mathbf{x}_i - b \leq \varepsilon + \xi_i \\ \mathbf{w}^T \mathbf{x}_i + b - y_i \leq \varepsilon + \xi_i^* \end{cases}$ untuk $i = 1, 2, \dots, n$
 $\xi_i, \xi_i^* \geq 0$

$C > 0$ merupakan parameter *cost* yang mengendalikan *trade-off* antara pinalti variabel *slack* dan *margin*. Secara umum nilai C mulai dari 0,001 hingga 1000 (Rodríguez-Pérez & Bajorath, 2022). *Loss function* merupakan fungsi yang menunjukkan hubungan antara *error* dengan bagaimana *error* ini diberi pinalti dan ε -insensitive loss function merupakan jenis *loss function* yang dapat digunakan untuk data *nonlinear* dan memungkinkan serangkaian *support vector* akan diperoleh. Formulasi ε -insensitive loss function sebagai berikut (Safira et al., 2023):

$$L_\varepsilon = \begin{cases} 0, \text{ untuk } |f(\mathbf{x}) - y_i| < \varepsilon \\ |f(\mathbf{x}) - y_i| - \varepsilon, \text{ untuk yang lain} \end{cases} \quad (4)$$

Penyelesaian optimasi untuk persamaan (3) dapat lebih mudah diselesaikan dengan fungsi *Lagrange*. Formula fungsi *lagrange multiplier* sebagai solusi optimal untuk persamaan (4) dengan pembatas pada persamaan (5) adalah sebagai berikut:

$$L = \left(\frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \right) - \left(\sum_{i=1}^n \alpha_i (\varepsilon + \xi_i - y_i + \langle \mathbf{w}^T \mathbf{x}_i \rangle + b) \right) - \left(\sum_{i=1}^n \alpha_i^* (\varepsilon + \xi_i^* + y_i - \langle \mathbf{w}^T \mathbf{x}_i \rangle - b) \right) - \left(\sum_{i=1}^n (\eta_i \xi_i + \eta_i^* \xi_i^*) \right) \quad (5)$$

dengan L adalah fungsi *lagrange*, $\alpha_i, \alpha_i^*, \eta_i, \eta_i^* \geq 0$ adalah koefisien *lagrange multiplier*.

Fungsi *lagrange multiplier* pada persamaan (6) digunakan untuk mendapatkan solusi yang optimal yang dapat dihitung dengan melakukan turunan parsial L terhadap $\mathbf{w}, b, \xi_i, \xi_i^*$ sekaligus L terhadap α_i , sehingga menghasilkan persamaan dual untuk masalah optimasi SVR sebagai berikut (Artini et al., 2024):

$$L_d = \frac{1}{2} \left(\sum_{j=1}^n \sum_{i=1}^n (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) \langle \mathbf{x}_i \cdot \mathbf{x}_j \rangle \right) - \varepsilon \sum_{i=1}^n (\alpha_i + \alpha_i^*) + \sum_{i=1}^n y_i (\alpha_i - \alpha_i^*) \quad (6)$$

dengan $\mathbf{x}_i \cdot \mathbf{x}_j$ merupakan *dot product* dari dua data dan koefisien *lagrange* α_i dan α_i^* memiliki batasan $\sum_{i,j=1}^n (\alpha_i - \alpha_i^*) = 0$, $0 \leq \alpha_i \leq C$ dan $0 \leq \alpha_i^* \leq C$, $i = 1, 2, \dots, n$. Penyelesaian untuk mendapatkan koefisien *lagrange* dapat dilakukan dengan menggunakan bantuan *software RStudio*.

Persamaan dual, yaitu bentuk turunan dari masalah optimisasi utama (*primal*) yang diperoleh melalui metode pengganda Lagrange, memuat variabel utama sebagai solusi optimal parameter \mathbf{w} dalam bentuk

koefisien Lagrange yaitu α_i dan α_i^* . Dengan demikian, nilai \mathbf{w} pada Persamaan (1) dapat ditulis kembali pada Persamaan (7) (Ishlah et al., 2023).

$$\mathbf{w} = \sum_{i=1}^n (\alpha_i - \alpha_i^*) \mathbf{x}_i \quad (7)$$

sehingga dihasilkan fungsi regresi SVR secara *eksplisit* yang dirumuskan sebagai berikut:

$$f(\mathbf{x}) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) (\mathbf{x}_i \cdot \mathbf{x}) + b \quad (8)$$

dengan α_i dan α_i^* adalah koefisien *lagrange*, $(\mathbf{x}_i \cdot \mathbf{x})$ adalah *dot product* yang menghitung hubungan antara titik data *training* \mathbf{x}_i dan titik input *testing* \mathbf{x} yang ingin diprediksi dan b adalah bias. Solusi optimal untuk b adalah menggunakan persamaan *Karush-Kuhn-Tucker* (KKT) sebagai berikut:

$$b = y_i - \mathbf{w}^T \mathbf{x}_i - \varepsilon, \text{ untuk } 0 \leq \alpha_i \leq C \quad (9)$$

dan

$$b = y_i - \mathbf{w}^T \mathbf{x}_i + \varepsilon, \text{ untuk } 0 \leq \alpha_i^* \leq C \quad (10)$$

Persamaan (8) merupakan persamaan yang digunakan untuk menghitung persamaan secara *linear* SVR dapat digunakan untuk kasus *nonlinear* dengan pendekatan alternatif melalui pemetaan data \mathbf{x} dari input *space* ke *feature space* dengan dimensi yang lebih tinggi melalui suatu fungsi φ sehingga $\varphi: \mathbf{x} \rightarrow \varphi(\mathbf{x})$. Oleh karena itu, fungsi regresi SVR *nonlinear* ditulis sebagai berikut (Artini et al., 2024):

$$f(\mathbf{x}) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) \varphi(\mathbf{x}_i) \varphi(\mathbf{x}) + b \quad (11)$$

transformasi φ pada umumnya tidak diketahui dan sulit dipahami, sehingga masalah ini dapat diatasi dengan fungsi *kernel* yaitu:

$$K(\mathbf{x}_i, \mathbf{x}) = \varphi(\mathbf{x}_i) \varphi(\mathbf{x}) \quad (12)$$

maka didapatkan persamaan akhir SVR *nonlinear* sebagai berikut:

$$f(\mathbf{x}) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) K(\mathbf{x}_i, \mathbf{x}) + b \quad (13)$$

dengan $K(\mathbf{x}_i, \mathbf{x})$ adalah fungsi *kernel* yang mengukur kesamaan antara data *training* \mathbf{x}_i dan data *testing* \mathbf{x} untuk proses prediksi, n adalah jumlah data *training* yang termasuk dalam *support vector*, yaitu data yang memiliki nilai α_i dan $\alpha_i^* \neq 0$.

2.2 Fungsi Kernel

Kinerja SVR dipengaruhi oleh jenis fungsi *kernel* dan pengaturan parameter *kernel* (Isnaeni et al., 2022). Fungsi ini digunakan untuk menghitung tingkat kemiripan (*similarity*) antara pasangan data, khususnya antar data *training* saat proses pelatihan model. Hasil dari perhitungan ini membentuk suatu matriks *kernel*, yaitu matriks simetris yang merepresentasikan hubungan antar seluruh data *training* dan menjadi dasar dalam menyusun model regresi (Ngu et al., 2024). Selama proses pelatihan, fungsi *kernel* dievaluasi terhadap seluruh pasangan titik data dalam data *training*, yang dituliskan sebagai $K(\mathbf{x}_i, \mathbf{x}_j)$. Sementara itu, saat model digunakan untuk melakukan prediksi terhadap data *testing*, fungsi *kernel* dievaluasi antara setiap titik data pelatihan \mathbf{x}_i dengan data input *testing* \mathbf{x} , yang dituliskan sebagai $K(\mathbf{x}_i, \mathbf{x})$. Beberapa jenis *kernel* yang digunakan pada metode SVR sebagai berikut (Meyer et al., 2024):

1. *Kernel Linear*

$$K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i \cdot \mathbf{x}_j \quad (14)$$

2. *Kernel Polynomial*

$$K(\mathbf{x}_i, \mathbf{x}_j) = (\gamma(\mathbf{x}_i \cdot \mathbf{x}_j) + c)^d \quad (15)$$

3. *Kernel Radial Basis Function (RBF)*

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2), \gamma > 0 \quad (16)$$

dengan $\mathbf{x}_i \cdot \mathbf{x}_j$ adalah *dot product* antara pasangan data input *training* saat pembentukan model, γ, c , dan d adalah parameter *kernel*. γ adalah parameter yang menentukan seberapa besar pengaruh setiap titik data dalam membuat pola prediksi. Parameter ini diperlukan untuk semua *kernel* kecuali *linear* dengan *default packages* **e1071** adalah $1/k$, dengan k adalah jumlah variabel prediktor. Parameter c adalah konstanta tambahan dalam fungsi *kernel polynomial* yang berperan dalam mengatur pergeseran (*offset*) dan mempengaruhi kontribusi hasil *dot product* $(\mathbf{x}_i \cdot \mathbf{x}_j)$. Nilai *default* c dalam *packages* **e1071** adalah 0 yang berarti fungsi *kernel* hanya bergantung pada *dot product* antara input. Sementara d (*degree*) adalah derajat fungsi *kernel polynomial* dengan nilai *default* 3 dalam *packages* **e1071** (Meyer et al., 2024). Pada penelitian ini,

kernel yang digunakan pada model SVR adalah *kernel polynomial*. Sehingga fungsi regresi SVR *nonlinear* dengan *kernel polynomial* adalah sebagai berikut:

$$f(\mathbf{x}) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) (\gamma(\mathbf{x}_i \cdot \mathbf{x}) + c)^d + b \quad (17)$$

2.3 Pembagian Data *Training* dan Data *Testing*

Perhitungan jumlah data *training* dan data *testing* dapat dihitung menggunakan persamaan berikut:

$$\sum \text{training} = \text{Proporsi data training} \times n \quad (18)$$

$$\sum \text{testing} = n - \text{jumlah data training} \quad (19)$$

dengan n adalah jumlah keseluruhan data.

2.4 Normalisasi dan Denormalisasi

Normalisasi memiliki tujuan untuk mendapatkan data dengan ukuran yang lebih kecil yang dapat mewakili data yang asli tanpa kehilangan karakteristiknya sendiri sedangkan denormalisasi adalah proses pengembalian data ke nilai awal yang sebelumnya telah dilakukan proses normalisasi untuk mendapatkan data asli. Salah satu metode normalisasi data yang umum digunakan adalah *Min-Max Normalization*. Metode *Min-Max Normalization* adalah metode normalisasi dengan melakukan transformasi *linear* terhadap data asli sehingga menghasilkan keseimbangan nilai perbandingan antar data saat sebelum dan sesudah proses. Metode ini merubah rentang nilai data menjadi antara 0 dan 1. Persamaan metode *Min-Max Normalization* adalah sebagai berikut.

$$x_{ig}^* = \frac{x_{ig} - \min(x_g)}{\max(x_g) - \min(x_g)} \quad (20)$$

Untuk mengembalikan nilai normalisasi yang merupakan hasil prediksi ke data asli atau disebut denormalisasi data adalah sebagai berikut:

$$x_{ig} = x_{ig}^* (\max(x_g) - \min(x_g)) + \min(x_g) \quad (21)$$

dengan x_{ig}^* adalah nilai hasil normalisasi data ke- i pada variabel yang dinormalisasi ke- g dan x_{ig} adalah nilai data aktual ke- i pada variabel ke- g (Artini et al., 2024).

2.5 Ukuran Kebaikan Model Prediksi

Salah satu metode yang digunakan untuk menghitung hasil evaluasi suatu model yang telah didapat dalam perhitungan adalah *Root Mean Square Error* (RMSE). RMSE merupakan besarnya tingkat *error* hasil prediksi, dimana semakin kecil (mendekati 0) nilai RMSE maka hasil prediksi akan semakin akurat. Nilai RMSE dapat dihitung dengan persamaan sebagai berikut (Artini et al., 2024):

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{Y}_i - Y_i)^2}{n}} \quad (22)$$

dimana n adalah jumlah keseluruhan data, \hat{Y}_i merupakan nilai prediksi pada periode i dan Y_i adalah nilai aktual pada periode i .

3. Metodologi

3.1 Data dan Variabel Penelitian

Data yang digunakan dalam penelitian ini adalah data curah hujan di Kabupaten Berau periode Januari 2014 sampai Desember 2023 yang diperoleh dari laman resmi Badan Pusat Statistik Provinsi Kalimantan Timur yaitu <https://kaltim.bps.go.id/>. Variabel yang digunakan terdiri atas variabel respon (Y) dan variabel prediktor (X_k), dimana $k = 1, 2, 3, 4, 5$ seperti yang disajikan pada Tabel 1.

Tabel 1. Variabel Penelitian

Simbol	Variabel	Defenisi Operasional	Skala Pengukuran	Satuan
Y	Curah Hujan	Curah hujan adalah volume air hujan yang turun di suatu daerah. Curah hujan 1 mm berarti air setinggi 1 mm terkumpul pada permukaan datar seluas 1 m^2 atau setara dengan 1 liter	Rasio	mm
X_1	Suhu	Ukuran seberapa dingin atau panas sebuah kondisi.	Interval	°C

X_3	Tekanan Udara	Berat sebuah kolom udara per satuan luas di atas sebuah titik, diukur dengan Barometer	Rasio	mb
X_4	Kecepatan Angin	Satuan yang mengukur kecepatan udara dari tekanan tinggi ke tekanan rendah	Rasio	knot
X_5	Penyinaran Matahari	Daya per satuan luas yang diterima matahari dalam bentuk radiasi elektromagnetik	Rasio	%

3.2 Populasi, Sampel, dan Teknik *Sampling*

Populasi dalam penelitian ini adalah curah hujan di Kabupaten Berau, sedangkan sampel yang digunakan adalah curah hujan di Kabupaten Berau dari Bulan Januari 2014 sampai dengan Desember 2023 sebanyak 120 bulan. Teknik *sampling* yang digunakan adalah *purposive sampling* yaitu teknik penentuan sampel dengan pertimbangan tertentu. Adapun yang menjadi pertimbangan penulis adalah ketersediaan data terbaru

3.3 Teknik Analisis Data

Teknik analisis data dalam penelitian ini adalah sebagai berikut:

1. Melakukan analisis statistika deskriptif terhadap variabel penelitian.
2. Melakukan normalisasi data menggunakan metode *Min-Max Normalization* pada Persamaan (20).
3. Melakukan pembagian data, menjadi data *training* dan *testing* dengan proporsi yang digunakan adalah 80%:20% menggunakan Persamaan (18) dan (19).
4. Membangun model SVR dengan fungsi *kernel polynomial* menggunakan parameter C dan d yang telah ditetapkan sebelumnya menggunakan Persamaan (17).
5. Memprediksi curah hujan di Kabupaten Berau menggunakan model SVR *kernel polynomial* yang telah diperoleh dari Langkah 4.
6. Melakukan denormalisasi data berdasarkan Persamaan (21).
7. Membuat grafik perbandingan data aktual dan prediksi pada model SVR dengan *kernel polynomial*.
8. Membandingkan ukuran kebaikan model prediksi curah hujan di Kabupaten Berau antara model SVR dengan fungsi *kernel polynomial* menggunakan akurasi RMSE pada Persamaan (22).
9. Menarik kesimpulan.

4. Hasil dan Pembahasan

4.1 Deskripsi Variabel Penelitian

Deskripsi variabel penelitian dilakukan dengan melakukan analisis statistika deskriptif. Hasil yang diperoleh disajikan pada Tabel 2.

Tabel 2. Deskripsi Variabel Penelitian

Variabel	Rata-rata	Simpangan Baku	Minimum	Maksimum
Curah Hujan (Y)	189,7	88	27	535
Suhu (X_1)	27,10	1,01	26	36,60
Kelembaban Udara (X_2)	86,81	2,85	78	93
Tekanan Udara (X_3)	1011,10	1,72	1007,45	1014,20
Kecepatan Angin (X_4)	3,12	1,08	1,14	5
Penyinaran Matahari (X_5)	38,76	21,17	2,46	81

Berdasarkan Tabel 2, dapat diketahui bahwa pada Tahun 2014 sampai 2023 Kabupaten Berau memiliki curah hujan sebesar 27 mm hingga 535 mm, dengan rata-rata curah hujan berada di 189,7 mm. Nilai simpangan baku sebesar 88 mm yang artinya data curah hujan di Kabupaten Berau menyebar sejauh 88 dari nilai rata-rata. Suhu udara sebesar 26°C hingga 36,60°C, dengan rata-rata suhu berada di 27,10°C. Nilai simpangan baku 1,01°C menunjukkan bahwa variasi suhu di kabupaten berau relatif kecil. Kelembaban udara tergolong lembab dimana kelembaban udara tertinggi mencapai 93 persen dan paling minim hanya berkisar 78 persen dengan rata-rata 86,81 persen. Nilai simpangan baku sebesar 2,85 persen menunjukkan variasi bulanan dari kelembaban udara tidak terlalu besar. Rata-rata tekanan udara sebesar 1011,10 mb, terendah sebesar 1007,45 mb dan tertinggi sebesar 1014,20 mb dengan standar deviasi sebesar 1,72 mb. Rata-rata kecepatan sebesar 3,12 knot dengan kecepatan terendah sebesar 1,14 knot dan kecepatan angin tertinggi sebesar 5 knot. Nilai simpangan baku sebesar 1,08 knot. Rata-rata penyinaran matahari sebesar 38,76% dengan penyinaran

terendah sebesar 2,46% dan penyinaran tertinggi sebesar 81%. Nilai simpangan baku sebesar 21,17% menunjukkan bahwa data penyinaran matahari menyebar cukup jauh dari nilai rata-rata.

4.2 Normalisasi Data

Normalisasi data dilakukan agar semua variabel berada dalam jangkauan yang sama sehingga proporsi pengaruh pada prediksi dapat seimbang. Penelitian ini menggunakan metode *Min-Max Normalization* pada Persamaan (20). Perhitungan dilakukan hingga pengamatan ke-120 untuk setiap variabel, sehingga diperoleh nilai dalam rentang setiap variabel adalah 0 – 1.

4.3 Pembagian Data *Training* dan Data *Testing*

Pada penelitian ini, proporsi pembagian data *training* dan *testing* menggunakan perbandingan 80%:20% menggunakan persamaan (18) dan (19). Berdasarkan perhitungan, didapatkan bahwa data *training* adalah data periode Januari 2014 – Desember 2021 sebanyak 96 bulan atau data ke-1 sampai data ke-96. Data *testing* adalah periode Januari 2022 – Desember 2023 sebanyak 24 bulan atau data ke-97 sampai data ke-120. Data *training* digunakan untuk membangun model SVR, sedangkan data *testing* digunakan untuk memprediksi dengan metode SVR berdasarkan model yang telah didapatkan sebelumnya.

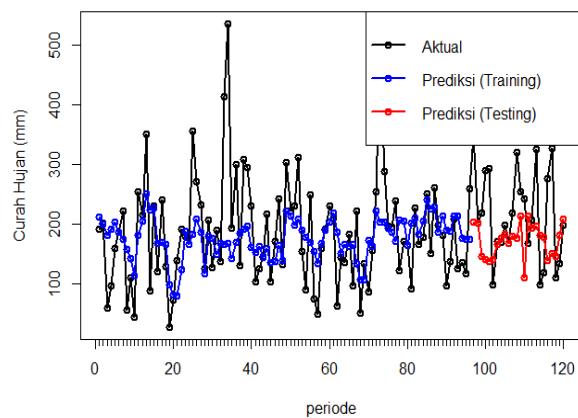
4.4 Prediksi Curah Hujan dengan Model SVR

Model SVR dapat diperoleh dengan bantuan *software RStudio* melalui *package* tambahan yaitu **e1071**. Proses dilakukan dengan menentukan nilai parameter *cost* (*C*) dan *degree* (*d*) terlebih dahulu. Nilai parameter *C* yang digunakan pada penelitian ini yaitu 0,5;1;2;4 dan *d* yaitu 1 dan 2. Selanjutnya membuat pasangan kombinasi nilai *C* dan *d*, dengan semua nilai *C* harus berpasangan dengan *d*, begitupun sebaliknya, sehingga akan terbentuk delapan pasang model SVR. Adapun nilai parameter *gamma* dan *c* untuk menghitung matriks *kernel polynomial* menggunakan nilai secara *default* dalam *package e1071* yaitu untuk *gamma*, karena terdapat lima variabel prediktor maka nilai *default*-nya adalah 0,2, dan *c* memiliki nilai *default* yaitu 0. Model yang terbentuk dari delapan kombinasi parameter *C* dan *d* yang digunakan disajikan pada Tabel 3.

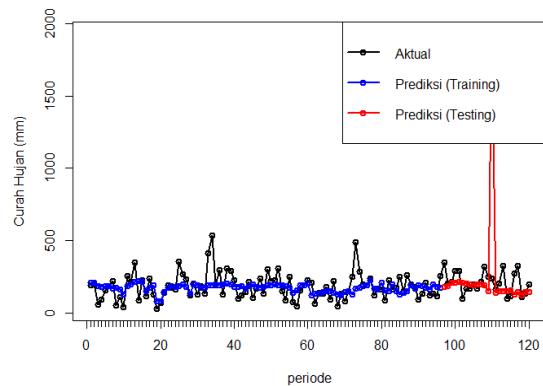
Tabel 3. Model SVR

Nomor	Parameter		Model SVR
	<i>C</i>	<i>d</i>	
1	0,5	1	$f(\mathbf{x}) = \sum_{i=1}^{84} (\alpha_i - \alpha_i^*)(0,2(\mathbf{x}_i \cdot \mathbf{x})) + 0,0819$
2	0,5	2	$f(\mathbf{x}) = \sum_{i=1}^{89} (\alpha_i - \alpha_i^*)(0,2(\mathbf{x}_i \cdot \mathbf{x}))^2 - 0,1478$
3	1	1	$f(\mathbf{x}) = \sum_{i=1}^{83} (\alpha_i - \alpha_i^*)(0,2(\mathbf{x}_i \cdot \mathbf{x})) + 0,079$
4	1	2	$f(\mathbf{x}) = \sum_{i=1}^{90} (\alpha_i - \alpha_i^*)(0,2(\mathbf{x}_i \cdot \mathbf{x}))^2 - 0,1834$
5	2	1	$f(\mathbf{x}) = \sum_{i=1}^{82} (\alpha_i - \alpha_i^*)(0,2(\mathbf{x}_i \cdot \mathbf{x})) + 0,0961$
6	2	2	$f(\mathbf{x}) = \sum_{i=1}^{89} (\alpha_i - \alpha_i^*)(0,2(\mathbf{x}_i \cdot \mathbf{x}))^2 - 0,231$
7	4	1	$f(\mathbf{x}) = \sum_{i=1}^{82} (\alpha_i - \alpha_i^*)(0,2(\mathbf{x}_i \cdot \mathbf{x}))^2 - 0,0948$

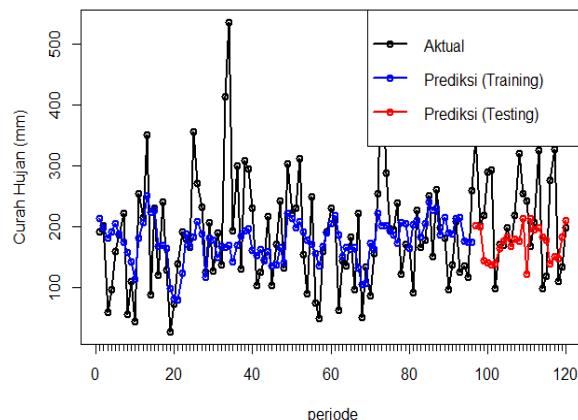
Setelah model terbentuk melalui proses *training*, selanjutnya dilakukan prediksi baik terhadap data *training* maupun data *testing*. Hasil prediksi yang diperoleh akan berada dalam skala normalisasi. Oleh karena itu, nilai hasil prediksi tersebut dikembalikan ke data asli yang disebut dengan denormalisasi data menggunakan Persamaan (21). Hasil prediksi dari model SVR pada Tabel 3 dapat dilihat pada Gambar 2, Gambar 3, Gambar 4, Gambar 5, Gambar 6, Gambar 7, Gambar 8 dan Gambar 9.



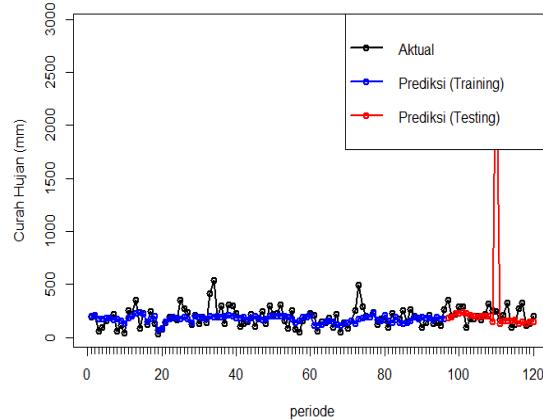
Gambar 2. Prediksi model SVR 1



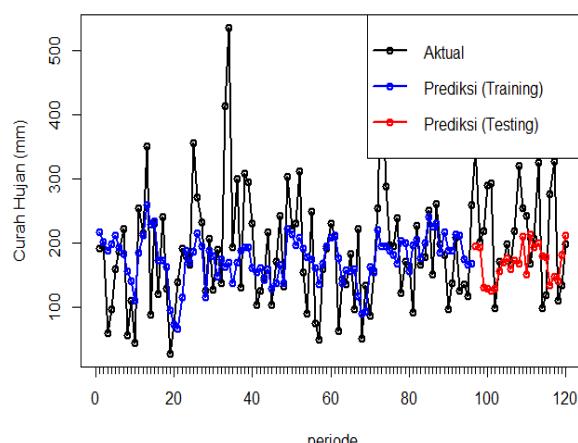
Gambar 3. Prediksi model SVR 2



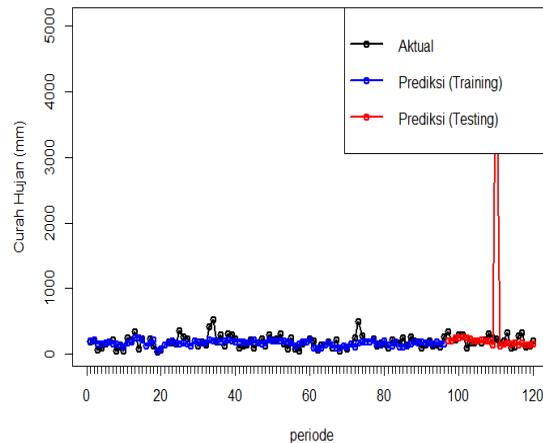
Gambar 4. Prediksi model SVR 3



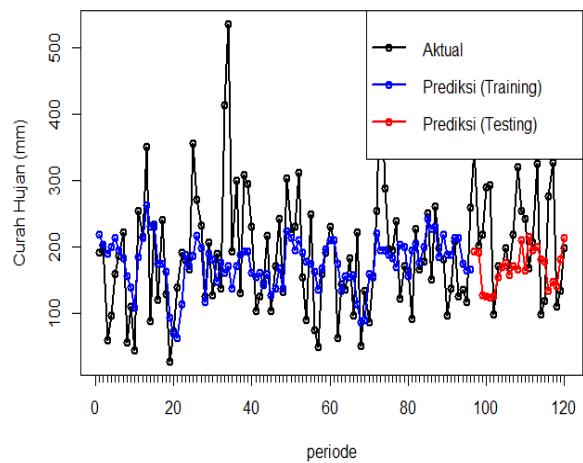
Gambar 5. Prediksi model SVR 4



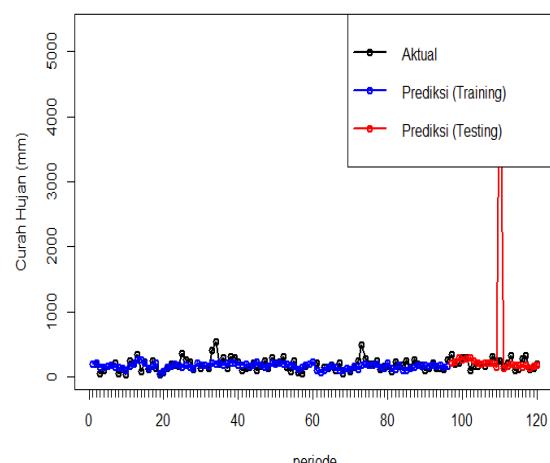
Gambar 6. Prediksi model SVR 5



Gambar 7. Prediksi model SVR 6



Gambar 8. Prediksi model SVR 7



Gambar 9. Prediksi model SVR 8

Visualisasi hasil prediksi curah hujan di Kabupaten Berau dengan delapan model SVR yang disajikan pada Gambar 2 hingga Gambar 9, menunjukkan hasil yang berbeda-beda saat pembentukan model (menggunakan data *training*) maupun pengujian model (menggunakan data *testing*). Model-model seperti SVR 1, SVR 3, SVR 5, dan SVR 7 (Gambar 2, 4, 6, dan 8) memperlihatkan performa yang relatif baik, di mana prediksi pada data *training* dan *testing* cukup mendekati pola data aktual, meskipun pada beberapa titik, terutama saat terjadi lonjakan curah hujan yang ejstrem, model menghasilkan prediksi yang lebih rendah dari nilai aktual. Sebaliknya, model SVR 2, SVR 4, SVR 6, dan SVR 8 (Gambar 3, 5, 7 dan 9) mengalami lonjakan prediksi yang jauh berbeda dari data aktual seperti lonjakan curah hujan yang terlalu tinggi pada bagian data *testing*. Oleh karena itu, secara keseluruhan, hanya sebagian model SVR yang menunjukkan hasil prediksi yang relatif baik dalam memprediksi tren dan pola curah hujan di Kabupaten Berau periode 2014 sampai Desember 2023, sedangkan sisanya menghasilkan prediksi yang kurang baik, terutama pada data *testing*.

4.5 Pemilihan Model SVR Terbaik

Tabel 4 menunjukkan nilai RMSE untuk pemilihan model terbaik dari delapan pasang model yang digunakan.

Tabel 4. Pemilihan Model SVR Terbaik

Model SVR	Parameter		
	<i>C</i>	<i>d</i>	RMSE
1	0,5	1	0,1795
2	0,5	2	0,7004
3	1	1	0,1786
4	1	2	1,0951
5	2	1	0,1832
6	2	2	1,9487
7	4	1	0,1836
8	4	2	2,0619

Berdasarkan Tabel 4, diperoleh model SVR terbaik adalah model SVR 3 yang ditunjukkan dengan nilai RMSE terkecil. Hal ini menunjukkan bahwa model SVR terbaik dengan *kernel polynomial* untuk memprediksi curah hujan di Kabupaten Berau adalah dengan menggunakan parameter $C = 1$ dan $d = 1$. Nilai tingkat akurasi prediksi RMSE sebesar 0,1786 menunjukkan model SVR memiliki performa yang baik dalam memprediksi curah hujan di Kabupaten Berau.

5. Kesimpulan

Berdasarkan hasil analisis data dan pembahasan, diperoleh kesimpulan sebagai berikut:

1. Model SVR terbaik untuk memprediksi curah hujan di Kabupaten Berau berdasarkan proporsi data *training* dan *testing* 80%:20% adalah model SVR dengan fungsi *kernel polynomial*, dan nilai parameter $C = 1$ dan parameter $d = 1$.
2. Hasil prediksi curah hujan di Kabupaten Berau berdasarkan model SVR terbaik adalah baik yang ditunjukkan oleh nilai RMSE sebesar 0,1786.

6. Referensi

- Abdullah, M., & Said, S. (2025). Performance evaluation of machine learning regression models for rainfall prediction. *Iranian Journal of Science and Technology, Transactions of Civil Engineering*, 49(4), 4231–4250.
- Artini, N. P. S. Y., Sumarjaya, I. W., & Nilakusmawati, D. P. E. (2024). Penerapan Metode Support Vector Regression (SVR) Dengan Algoritma Grid Search Dalam Peramalan Harga Saham.
- Badan Penanggulangan Bencana Daerah Kabupaten Berau. (2021). Dokumen Peta Risiko Bencana Kabupaten Berau Provinsi Kalimantan Timur Tahun 2020 - 2024.
- Badan Pusat Statistik. (2023). Provinsi Kalimantan Timur Dalam Angka 2023. Samarinda: Badan Pusat Statistik Provinsi Kalimantan Timur.
- Hermawan, A., Mangku, I. W., Ardana, N. K. K., & Sumarno, H. (2022). Analisis support vector regression dengan algoritma grid search untuk memprediksi harga saham. *MILANG Journal of Mathematics and Its Applications*, 18(1), 41–60.
- Ishlah, A. W., Sudarno, S., & Kartikasari, P. (2023). Implementasi gridSearchCV pada support vector regression (SVR) untuk peramalan harga saham. *Jurnal Gaussian*, 12(2), 276–286.
- Isnaeni, R., Sudarmin, S., & Rais, Z. (2022). Analisis Support Vector Regression (Svr) Dengan Kernel Radial Basis Function (Rbf) Untuk Memprediksi Laju Inflasi Di Indonesia. *VARIANSI: Journal of Statistics and Its Application on Teaching and Research*, 4(1), 30–38.
- Karondia, L. A., Fitrian, R., & Hizkia, H. (2022). Pemetaan Zonasi Kerawanan Banjir berbasis Sistem Informasi Geografis (Studi Kasus: Kabupaten Berau, Kalimantan Timur). *Geoid*, 18(1), 57–68.
- Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., Leisch, F., Chang, C. C., & Lin, C. C. (2024). Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien [R package e1071 version 1.7-16]. *CRAN Contrib. Packag*.
- Ngu, J. C. Y., Yeo, W. S., Thien, T. F., & Nandong, J. (2024). A comprehensive overview of the applications of kernel functions and data-driven models in regression and classification tasks in the context of software sensors. *Applied Soft Computing*, 164, 111975.
- Nurhalizah, R. S., Ardianto, R., & Purwono, P. (2024). Analisis Supervised dan Unsupervised Learning pada Machine Learning: Systematic Literature Review. *Jurnal Ilmu Komputer Dan Informatika*, 4(1), 61–72.
- Purwoko, C. F. F., Sediono, S., Saifudin, T., & Mardianto, M. F. F. (2023). Prediksi Harga Eksport Non Migas di Indonesia Berdasarkan Metode Estimator Deret Fourier dan Support Vector Regression. *Inferensi*, 6(1), 45–55.
- Rodríguez-Pérez, R., & Bajorath, J. (2022). Evolution of support vector machine and regression modeling in chemoinformatics and drug discovery. *Journal of Computer-Aided Molecular Design*, 36(5), 355–362.
- Safira, A. N., Warsito, B., & Rusgiyono, A. (2023). Analisis support vector regression (SVR) dengan algoritma grid search time series cross validation untuk prediksi jumlah kasus terkonfirmasi COVID-19 di Indonesia. *Jurnal Gaussian*, 11(4), 512–521.
- Siregar, N. A. (2022). Peramalan curah hujan di Kota Medan menggunakan metode support vector regression. *Journal of Informatics and Data Science (J-IDS)*, 1(1), 1–3.
- Sunarmi, N., Kumailia, E. N., Nurfaiza, N., Nikmah, A. K., Aisyah, H. N., Sriwahyuni, I., & Lailly, S. N. (2022). Analisis faktor unsur cuaca terhadap perubahan iklim di Kabupaten Pasuruan pada tahun 2021 dengan metode Principal Component Analysis. *Newton-Maxwell Journal of Physics*, 3(2), 56–64.
- Suyono, A. A., Kusrini, K., & Arief, M. R. (2022). Prediksi Indeks Harga Konsumen Komoditas Makanan di Kota Surabaya menggunakan Support Vector Regression. *METIK JURNAL (AKREDITASI SINTA 3)*, 6(1), 45–51.