

# Penerapan Metode *Ensemble Bagging-SVR* Untuk *Multivariate Time Series* Pada Peramalan Curah Hujan (Studi Kasus: Stasiun Meteorologi Perak I)

**Mochammad Rizki Aji Santoso<sup>1\*</sup>, Mula'ab<sup>2</sup>, Rika Yunitarini<sup>3</sup>**

<sup>1,2,3</sup> Teknik Informatika, Universitas Trunojoyo Madura, Indonesia

E-mail: <sup>1</sup>[200411100086@student.trunojoyo.ac.id](mailto:200411100086@student.trunojoyo.ac.id), <sup>2</sup>[mulaab@trunojoyo.ac.id](mailto:mulaab@trunojoyo.ac.id),

<sup>3</sup>[rika.yunitarini@trunojoyo.ac.id](mailto:rika.yunitarini@trunojoyo.ac.id)

**DOI:** <https://doi.org/10.52620/sainsdata.v4i1.294>

## ABSTRAK

Perubahan iklim menjadi salah satu isu paling kritis saat ini, dengan pengaruh yang dirasakan di berbagai negara, termasuk Indonesia. Fenomena ini memengaruhi berbagai aspek kehidupan, salah satunya adalah perubahan pola cuaca. Musim hujan sering ditandai dengan peningkatan intensitas curah hujan secara signifikan dalam jangka waktu tertentu. Cuaca yang tidak menentu dapat menyebabkan perubahan curah hujan yang drastis, berdampak pada sektor pertanian, transportasi, dan industri. Oleh karena itu, prediksi curah hujan menjadi penting untuk memahami pola hujan di masa depan. Penelitian ini menggunakan data iklim harian dari BMKG Stasiun Meteorologi Perak I Surabaya selama Januari 2018 hingga Desember 2024 dengan 2.557 data yang mencakup 4 parameter: curah hujan, temperatur, kelembapan, dan kecepatan angin. Tahapan penelitian meliputi pra pemrosesan data, seperti penanganan data hilang, identifikasi outlier, normalisasi, dan sliding window untuk membentuk data supervised. Metode Ensemble berbasis Support Vector Regression (SVR) digunakan untuk membangun model prediksi. Hasil evaluasi menunjukkan bahwa skenario 9, dengan kernel polynomial,  $C = 1$ ,  $\epsilon = 0,01$ ,  $\text{degree} = 2$ , dan 20 estimator, memberikan nilai MAE terbaik sebesar 0,02329. Hal ini menunjukkan bahwa pemilihan kernel dan hyperparameter yang tepat sangat berpengaruh terhadap akurasi model.

**Kata Kunci:** Curah Hujan, Metode *Ensemble*, Peramalan, Perubahan Iklim, *Support Vector Regression*.

## ABSTRACT

*Climate change has become one of the most critical issues today, with impacts felt in various countries, including Indonesia. This phenomenon affects various aspects of life, one of which is changes in weather patterns. The rainy season is often marked by a significant increase in rainfall intensity over a certain period. Unpredictable weather can cause drastic changes in rainfall, impacting agriculture, transportation, and industry. Therefore, rainfall prediction is essential to understand future rainfall patterns. This study uses daily climate data from the BMKG Perak I Surabaya Meteorological Station from January 2018 to December 2024 with 2,557 data covering 4 parameters: rainfall, temperature, humidity, and wind speed. The research stages include data pre-processing, such as handling missing data, outlier identification, normalization, and sliding window to form supervised data. An Ensemble-based Support Vector Regression (SVR) method is used to build a prediction model. The evaluation results show that scenario 9, with a polynomial kernel,  $C = 1$ ,  $\epsilon = 0.01$ ,  $\text{degree} = 2$ , and 20 estimators, provides the best MAE value of 0.02329. This indicates that the selection of the appropriate kernel and hyperparameters significantly influences the model's accuracy.*

**Keywords:** Climate Change, Ensemble Method, Forecasting, Rainfall, Support Vector Regression.



This work is licensed under a [Creative Commons Attribution 4.0 International License](#)

© 2025 Author (s)

## PENDAHULUAN

Perubahan iklim menjadi salah satu isu global saat ini. Dampaknya telah dirasakan di berbagai negara di belahan dunia, termasuk Indonesia. Fenomena ini telah mempengaruhi berbagai aspek kehidupan manusia, peningkatan suhu global, peningkatan permukaan air laut, perubahan pola angin termasuk pola cuaca dan distribusi curah hujan. Dampak perubahan iklim tidak hanya terbatas pada lingkungan, tetapi juga berdampak signifikan di sektor-sektor seperti kesehatan, pertanian, ekonomi, dan infrastruktur.

Pola cuaca juga sangat mempengaruhi berbagai aspek kehidupan sehari-hari, seperti pertanian, transportasi, dan industri. Dalam pertanian, perubahan iklim ekstrem membawa dampak besar bagi sektor pertanian. Para petani mengalami kesulitan dalam menentukan waktu penanaman yang tepat, sehingga mengalami gagal panen dan kerugian. Hal ini menyebabkan penurunan pendapatan petani, dan menjadikan sektor pertanian sangat rentan terhadap perubahan iklim [1]. Sementara dalam transportasi, Faktor keadaan cuaca misalnya hujan lebat, perbedaan suhu dan perbedaan kecepatan angin akan mempengaruhi pelayanan dan kelancaran jalur transportasi [2]. Di sisi lain, dalam industri, banyak perusahaan masih bergantung pada cuaca, seperti industri gula. Faktor iklim, khususnya curah hujan, sangat berperan dalam pertumbuhan dan hasil panen tebu. Hal ini akan mempengaruhi kadar gula pada nira tebu, dan pada akhirnya menentukan jumlah produksi gula yang dihasilkan [3]. Oleh karena itu, pemantauan kondisi cuaca, terutama curah hujan, menjadi sangat penting.

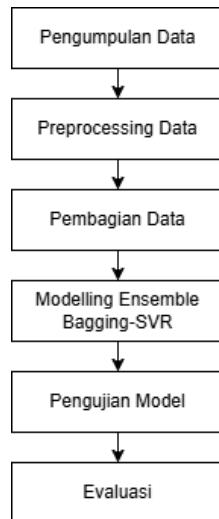
Hujan adalah butiran air atau es yang turun dari awan. Awan terbentuk dari kondensasi uap air yang terbawa angin dan tersebar di seluruh permukaan bumi. Butiran air yang cukup besar akan jatuh ke bumi sebagai hujan. Hujan merupakan bentuk endapan yang paling umum dijumpai, dan di Indonesia, curah hujan adalah istilah yang digunakan untuk menunjukkan jumlah endapan hujan [4]. Curah hujan tidak dapat dipastikan secara pasti, tetapi bisa diprediksi. Prediksi ini dilakukan dengan memanfaatkan data curah hujan di masa lalu untuk memperkirakan curah hujan di masa depan. Berbagai metode dapat dilakukan untuk memprediksi curah hujan, salah satunya menggunakan metode ensemble.

Di tengah tantangan ini, metode ensemble telah menjadi pendekatan yang populer dalam meningkatkan akurasi peramalan. Metode ensemble adalah teknik menggabungkan prediksi dari beberapa model atau teknik peramalan untuk memperkuat sebuah model sehingga dapat mencapai kinerja optimal saat digabungkan dengan model lainnya, demi memperoleh akurasi prediksi yang lebih baik [5]. Seperti penelitian yang dilakukan oleh Bayu Sugara dan Agus Subekti, hasil eksperimen menunjukkan bahwa metode ensemble terbukti meningkatkan ketepatan prediksi. Model SVM dengan kernel polinomial dan metode ensemble bagging mencapai nilai akurasi tertinggi, yaitu 91% [6]. Penelitian yang menggunakan neural network untuk prediksi harga listrik menggunakan ensemble bagging memiliki tingkat akurasi yang lebih unggul dari neural network biasa, yaitu memiliki nilai RMSE 10,513 dari yang awalnya sebesar 21,476 [7]. Penelitian lain Untuk memprediksi data runtun waktu univariat (laba Bank Syariah Indonesia), hasil penelitian menunjukkan bahwa ensemble stacking menggunakan algoritma Neural Network, Random Forest, SVM, Random Forest, dan General Linear Model memberikan prediksi yang paling akurat dengan RMSE sebesar 0,534 [5]. Metode ensemble tidak hanya meningkatkan akurasi sistem pengambilan keputusan, tetapi juga mengatasi berbagai tantangan seperti seleksi fitur, estimasi kepercayaan, penanganan data yang hilang, koreksi kesalahan, dan penanganan kelas data yang tidak seimbang. Ini menjadikan metode ensemble sangat efektif dan bermanfaat di berbagai domain dan aplikasi dunia nyata [8].

Metode Bagging (Bootstrap Aggregating), sebuah teknik ensemble dalam supervised learning, terbukti efektif dalam meningkatkan akurasi prediksi. Bagging bekerja dengan cara menghasilkan beberapa sampel data yang baru dari data pelatihan asli untuk memperoleh data pelatihan yang optimal dalam pembuatan model [7]. Support Vector Regression (SVR) adalah model regresi yang berasal dari SVM. Digunakan untuk menyelesaikan masalah non-linear. SVR berhasil diterapkan pada berbagai masalah prediksi time series [9]. Oleh karena itu, penelitian ini menerapkan Support Vector Regression dengan tambahan Bootstrap Aggregating (Bagging) untuk meramalkan curah hujan dengan harapan dapat meningkatkan performa prediksi yang lebih baik lagi.

## METODE

Tahapan pada penelitian ini akan dijelaskan pada gambar 1 berikut:

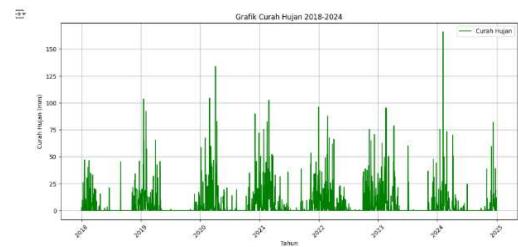


**Gambar 1.** Alur Penelitian

Dalam penelitian ini, langkah awal melibatkan pengumpulan data, yang kemudian diproses untuk memastikan kualitasnya sebelum dianalisis. Selanjutnya, data dibagi menjadi dua bagian: data latih untuk membangun model dan data uji untuk menilai kinerjanya. Model dikembangkan menggunakan data latih, sementara evaluasi dilakukan dengan data uji. Hasil evaluasi pada tahap pengujian digunakan untuk menentukan model yang paling optimal.

### Pengumpulan Data

Data yang digunakan Data yang digunakan berasal dari hasil observasi dari Badan Meteorologi, Klimatologi dan Geofisika (BMKG) Stasiun Meteorologi Perak I. yang diambil dari website <https://dataonline.bmkg.go.id/dataonline-home>. Data harian selama lima tahun terakhir mulai Januari 2018 hingga Desember 2024 digunakan untuk penelitian ini. Untuk jumlah total data hingga saat ini berjumlah 2.557 data. Parameter yang digunakan pada penelitian adalah 4 parameter yakni: tingkat curah hujan dengan satuan millimeter (mm), titik kelembapan rata-rata dengan satuan persen (%), temperatur rata-rata dengan satuan derajat Celcius (°C), dan kecepatan angin rata-rata dengan satuan meter per sekon (m/s) [10]. Data yang diperoleh dapat dilihat pada Gambar 2 sebagai berikut.



**Gambar 2.** Data Curah Hujan

### Preprocessing Data

Preprocessing adalah langkah paling awal sebelum data diproses. Pada tahap ini, data dianalisis secara menyeluruh untuk mempersiapkannya sebelum digunakan dalam model. Analisis data mengurangi kompleksitas dengan menghapus data yang tidak penting dan tidak relevan, menghindari noise, dan menemukan dan menghapus komponen yang tidak diperlukan [11].

## Interpolasi Linier

Pada proses ini akan dilakukan interpolasi data untuk menghilangkan missing value dengan sebuah nilai yang didapatkan dari dataset. Metode yang digunakan adalah metode untuk memperkirakan nilai pada titik tengah dengan menggunakan fungsi  $F(X)$  yang didasarkan pada dua nilai yang diketahui, yaitu nilai sebelumnya  $F(X_0)$  dan nilai sesudahnya  $F(X_1)$  [12]. Rumus interpolasi dapat dilihat pada persamaan (1) berikut:

$$f_1(X) = f(X_0) + \frac{f(X_1) - f(X_0)}{(X_1 - X_0)} (X - X_0) \quad (1)$$

## Identifikasi Outlier Dengan Z-Score

Menurut Hawkins (1980), outlier didefinisikan sebagai pengamatan yang menyimpang secara signifikan dari pengamatan lainnya. Keberadaan outlier ini dapat mempengaruhi kesimpulan atau keputusan dalam sebuah penelitian [13]. Deteksi data outlier dapat dilakukan dengan melakukan standarisasi data (Z-Score). Z-Score adalah teknik statistik yang dapat diterapkan pada big data. Z-Score, yang juga dikenal sebagai nilai baku atau nilai standar, sering digunakan dalam data mining untuk mengidentifikasi data yang merupakan outlier. Z-Score dapat dihitung menggunakan rumus berikut [14]:

$$z = \frac{x - \mu}{\sigma} \quad (2)$$

## Normalisasi Data

Normalisasi data merupakan metode untuk mengubah data agar memiliki bobot yang seimbang. Tujuan dari normalisasi adalah untuk mendapatkan data dengan ukuran lebih kecil yang tetap mampu merepresentasikan data asli tanpa menghilangkan karakteristik utamanya. Pada proses ini dilakukan normalisasi min-max untuk menstabilkan data. Metode yang dilakukan dengan merubah data menjadi bentuk skala dengan rentang nilai antara 0 hingga 1. Rumus normalisasi dapat dilihat pada persamaan 3 [15].

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (3)$$

## Penentuan Input dan Output

Sebelum memasuki tahap pembagian data training dan testing dan diproses dengan Bagging-SVR, dilakukan penentuan input dan output kedalam bentuk sliding window. Penentuan ini menggunakan *Partial Autocorrelation Function* (PACF) untuk menentukan fitur-lag yang akan digunakan. Fitur-lag (data  $t-1, t-2, \dots$ ) ditambahkan untuk membuat dataset menjadi bentuk *supervised*. Fungsi PACF mengukur hubungan antara  $X_t$  dan  $X_{t-k}$  dengan menghilangkan pengaruh lag sebelumnya. Untuk menentukan lag terbaik ( $k$ -terbaik), biasanya digunakan plot PACF yang menampilkan koefisien PACF pada berbagai lag. Plot ini membantu dalam memahami keterkaitan variabel pada waktu tertentu dengan nilai sebelumnya setelah menghilangkan efek lag lainnya. Nilai koefisien PACF sampel pada lag ke-  $k$  dihitung menggunakan Persamaan 4 [16].

$$\varphi_{kk} = \frac{r_k - \sum_{j=1}^{k-1} \varphi_{k-1,j} r_{k-j}}{1 - \sum_{j=1}^{k-1} \varphi_{k-1,j} r_j} \quad (4)$$

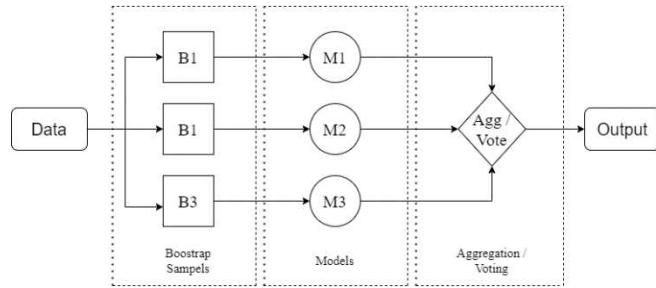
Fitur lag ini digunakan untuk memodelkan hubungan antara curah hujan saat ini dengan data sebelumnya. Namun, tidak hanya curah hujan saja yang ditambahkan berdasarkan hasil sliding window, tetapi juga variabel-variabel lain seperti kelembapan, temperatur, dan kecepatan angin. Hal ini dilakukan karena kelembapan, temperatur, dan kecepatan angin dianggap sebagai variabel eksogen dalam model. Variabel eksogen adalah variabel yang memengaruhi variabel target (dalam hal ini, curah hujan) tetapi tidak dipengaruhi oleh curah hujan itu sendiri.

## Pembagian Data

Pada penelitian ini, dataset yang digunakan adalah data curah hujan yang telah dikumpulkan untuk keperluan analisis dan pengujian model, dataset ini dibagi menjadi dua bagian, yaitu data latih (training set) dan data uji (test set). Pembagian ini dilakukan dengan rasio 80:20, di mana 80% dari data digunakan sebagai data latih dan 20% sisanya sebagai data uji.

## Pembentukan Model Bagging-SVR

Bagging (*Bootstrap Aggregating*) adalah teknik yang digunakan dalam regresi dan klasifikasi untuk meningkatkan stabilitas model. Dalam regresi, prediksi akhir diperoleh dengan mengambil rata-rata hasil dari beberapa model. Metode ini efektif dalam mengurangi variabilitas akibat perubahan kecil pada data latih [17]. *Support Vector Regression* (SVR) merupakan modifikasi dari SVM yang diterapkan pada regresi, di mana outputnya berupa nilai kontinu, berbeda dengan SVM yang digunakan untuk klasifikasi [15]. Dalam penelitian ini, SVR akan digunakan sebagai model dasar dalam pengembangan model. Parameter utama seperti  $C$ ,  $\epsilon$ ,  $\gamma$ ,  $d$ , dan jenis *kernel* pada SVR, serta *n\_estimator* pada Bagging, akan dioptimalkan guna meningkatkan kinerja model. Arsitektur Bagging-SVR dapat dilihat pada Gambar 3.



Gambar 3. Arsitektur Bagging-SVR

## Pengujian Model

Pengujian model dilakukan untuk menilai performa Bagging-SVR dalam peramalan curah hujan. Model yang telah dilatih menggunakan data latih akan diuji dengan data uji untuk mengamati hasil prediksi yang dihasilkan. Parameter utama pada SVR dan Bagging yang telah dioptimalkan akan diterapkan dalam proses ini guna memastikan model bekerja secara optimal. Selain itu, pengujian dilakukan untuk melihat bagaimana model menangkap pola data sebelum dilakukan tahap evaluasi lebih lanjut.

## Evaluasi

Model yang telah berhasil dibuat pada tahap sebelumnya kemudian akan dievaluasi untuk mengukur seberapa efektif prediksi yang dihasilkan. Dalam penelitian ini, evaluasi model yang digunakan meliputi beberapa metode yaitu RMSE dan MAE. *Root Mean Square Error* (RMSE) adalah metrik yang digunakan untuk mengukur tingkat kesalahan antara nilai aktual dan nilai prediksi. RMSE memberikan gambaran tentang seberapa besar perbedaan hasil prediksi terhadap nilai sebenarnya. Semakin kecil nilai RMSE (mendekati 0), semakin tinggi akurasi prediksi model [18]. Rumus RMSE dapat dilihat pada persamaan 5.

$$RMSE = \sqrt{\sum_{i=1}^n \frac{(\hat{y}_i - y_i)^2}{n}} \quad (5)$$

*Mean Absolute Error* (MAE) adalah metrik yang digunakan untuk mengevaluasi akurasi metode peramalan dengan menghitung rata-rata dari kesalahan absolut antara hasil prediksi dan nilai aktual. MAE menghitung jumlah total kesalahan absolut dan kemudian di rata-rata [18]. Rumus MAE dapat dilihat pada persamaan 6.

$$MAE = \frac{1}{n} \sum_{i=1}^n |Y_i - F(X_i)| \quad (6)$$

## Skenario Pengujian

Dalam pengujian ini, data dibagi dengan rasio 80:20. SVR digunakan dengan berbagai kernel seperti *Linear*, *Polynomial*, dan *RBF*, masing-masing dengan parameter tertentu. Pengujian dilakukan dalam sembilan skenario dengan jumlah *bootstrap samples* yang berbeda, namun menggunakan rasio pembagian data yang sama. Detail percobaan disajikan pada Tabel 1.

**Tabel 1.** Skenario Pengujian

Percobaan Ke-	Pembagian Data	Bootstraps Samples	Kernel
1	80% Training,	5	Linear
2	20% Testing	10	
3		20	
4		5	RBF
5		10	
6		20	
7		5	Polynomial
8		10	
9		20	

## HASIL DAN PEMBAHASAN

Dalam penelitian ini, pengujian dilakukan menggunakan metode *grid search* untuk mengevaluasi berbagai kombinasi parameter. Tujuan dari proses ini adalah menemukan konfigurasi parameter optimal guna meningkatkan akurasi model dalam peramalan curah hujan. Hasil prediksi dengan sembilan skenario pengujian disajikan pada Tabel 2.

**Tabel 2.** Skenario Hasil Dari Grid Search

Skenario Ke-	Kernel	Parameter					Evaluasi	
		C	$\epsilon$	$\gamma$	d	n_estimators	RMSE	MAE
1	Linear	0,01	0,01	-	-	5	0,072511	0,024008
2		0,01	0,01	-	-	10	0,072542	0,024023
3		0,01	0,01	-	-	20	0,072530	0,023973
4	RBF	1	0,01	0,01	-	5	<b>0,072318</b>	0,024013
5		1	0,01	0,01	-	10	0,072327	0,023885
6		1	0,01	0,01	-	20	0,072328	0,023846
7	Poly	1	0,01	-	2	5	0,073119	0,023643
8		1	0,01	-	2	10	0,072874	0,023381
9		1	0,01	-	2	20	0,072695	<b>0,023299</b>

Hasil pengujian menunjukkan bahwa performa *Ensemble Bagging-SVR* sangat dipengaruhi oleh jenis *kernel*, *hyperparameter*, dan jumlah estimator. *Grid search* digunakan untuk menentukan kombinasi optimal dari parameter C, epsilon ( $\epsilon$ ), gamma ( $\gamma$ ), dan degree (d), guna meminimalkan kesalahan prediksi. Setiap kombinasi parameter diuji dalam sembilan skenario dengan jumlah *bootstrap samples* yang berbeda.

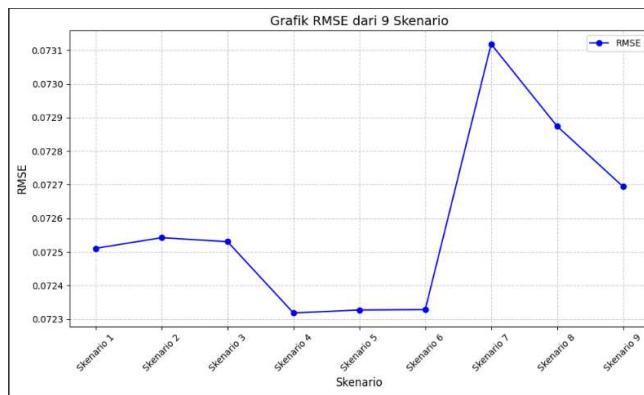
*Kernel RBF* memberikan hasil terbaik dalam skenario 4 hingga 6. Pada skenario 4, kombinasi parameter C = 1,  $\epsilon$  = 0,01,  $\gamma$  = 0,01, dan 5 estimator menghasilkan RMSE terendah sebesar 0,072318. Penambahan jumlah estimator pada skenario 5 dan 6 menjadi 10 dan 20 hanya memberikan sedikit perubahan dengan RMSE masing-masing sebesar 0,072327 dan 0,072328. Hal ini menunjukkan bahwa jumlah estimator yang lebih besar tidak selalu meningkatkan akurasi secara signifikan, tetapi berkontribusi pada stabilitas prediksi.

*Kernel Polynomial* juga memberikan hasil yang kompetitif, terutama pada skenario 9, dengan RMSE sebesar 0,072695 dan MAE terendah 0,023299. Kombinasi hyperparameter C = 1,  $\epsilon$  = 0,01, dan degree polynomial d = 2 menunjukkan bahwa *kernel Polynomial* dapat menangkap pola *non-linear* dengan cukup baik. Namun, pada

skenario dengan jumlah estimator lebih kecil (skenario 7 dan 8), performa *kernel Polynomial* masih kalah dibandingkan *kernel RBF*, meskipun tetap dalam batas yang dapat diterima.

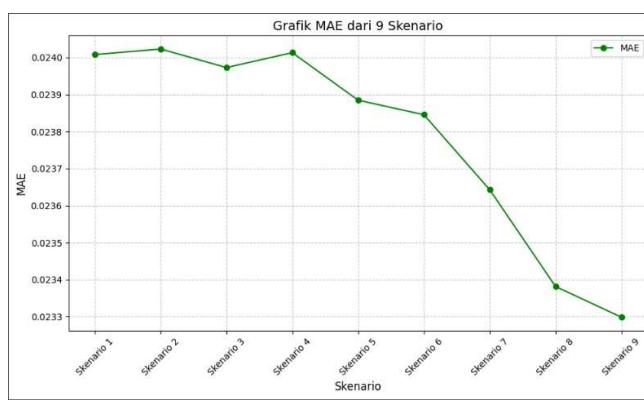
Sebaliknya, *kernel Linear* memiliki performa yang lebih rendah dibandingkan *kernel non-linear* lainnya. RMSE tertinggi sebesar 0,072542 tercatat pada skenario 2 dengan 10 estimator, sedangkan nilai MAE berkisar antara 0,024008 hingga 0,024023. Hal ini mengindikasikan bahwa *kernel Linear* kurang mampu menangkap pola kompleks dalam dataset curah hujan yang bersifat *non-linear*.

Selain itu, analisis juga menunjukkan bahwa peningkatan jumlah estimator dari 5 menjadi 10 dan 20 pada *kernel RBF* dan *Polynomial* tidak memberikan perubahan signifikan pada nilai RMSE maupun MAE. Model sudah mencapai kestabilan performa dengan jumlah estimator yang lebih kecil, sehingga menambah lebih banyak estimator hanya memberikan kontribusi marginal terhadap akurasi, tetapi dapat meningkatkan waktu komputasi. Oleh karena itu, pemilihan jumlah estimator yang optimal perlu mempertimbangkan *trade-off* antara akurasi dan efisiensi komputasi.



**Gambar 4.** Grafik Hasil RMSE

Pada grafik RMSE, terlihat bahwa nilai RMSE cenderung stabil pada skenario 1 hingga skenario 6, dengan penurunan yang cukup signifikan pada skenario 4. Namun, pada skenario 7, yang menggunakan kernel Polynomial dengan parameter  $d = 2$ , RMSE meningkat cukup tajam hingga mencapai nilai 0.073119. Hal ini mengindikasikan bahwa penggunaan kernel Polynomial pada skenario ini kurang optimal dalam menangani data dibandingkan kernel RBF. Penurunan kembali terjadi pada skenario 8 dan 9, dengan skenario 9 memiliki RMSE yang lebih rendah dibandingkan skenario 7 namun tetap lebih tinggi dibandingkan skenario 4-6.



**Gambar 5.** Grafik Hasil RMSE

Grafik MAE menunjukkan pola yang sedikit berbeda. Nilai MAE secara konsisten menurun dari skenario 1 hingga skenario 9, dengan skenario 9 menunjukkan nilai MAE paling rendah, yaitu 0.023299. Penurunan ini menandakan bahwa model pada skenario 9 memiliki performa terbaik dalam memprediksi nilai curah hujan dengan deviasi rata-rata yang lebih kecil dari nilai sebenarnya.

Hasil pengujian ini menegaskan bahwa pemilihan kernel yang tepat, pengaturan hyperparameter yang optimal melalui grid search, dan konfigurasi jumlah estimator memberikan pengaruh besar terhadap performa model.

Kernel RBF secara umum memiliki performa paling unggul dalam memprediksi curah hujan, terutama dalam menangkap pola non-linear yang dominan. Sementara itu, kernel Polynomial menunjukkan potensi baik dalam meminimalkan kesalahan absolut (MAE).

## SIMPULAN

Penelitian ini menunjukkan bahwa metode Support Vector Regression (SVR) dengan pendekatan ensemble bagging memberikan hasil yang baik dalam peramalan curah hujan. Dari sembilan skenario pengujian, performa model sangat dipengaruhi oleh jenis kernel, jumlah estimator, dan kombinasi hyperparameter.

Kernel RBF memberikan hasil terbaik pada skenario 4 dengan RMSE sebesar 0,072318 dan MAE sebesar 0,023846, menggunakan kombinasi hyperparameter  $C = 1$ ,  $\text{epsilon} = 0,01$ ,  $\gamma = 0,01$ , dan jumlah estimator 5. Sementara itu, kernel Polynomial menunjukkan performa terbaik pada skenario 9 dengan RMSE sebesar 0,072695 dan MAE 0,023299 dengan kombinasi  $C = 1$ ,  $\text{epsilon} = 0,01$ ,  $\text{degree} = 2$ , dan jumlah estimator 20. Kernel Linear memiliki performa yang lebih rendah dibandingkan kernel non-linear, dengan hasil terbaik pada skenario 1 (RMSE 0,072511, MAE 0,024008).

Meskipun jumlah estimator yang lebih besar dapat meningkatkan stabilitas model, hasil pengujian menunjukkan bahwa peningkatan jumlah estimator dari 5 ke 10 atau 20 tidak selalu memberikan peningkatan akurasi yang signifikan. Selain itu, tuning hyperparameter seperti nilai  $C$ ,  $\text{epsilon}$ ,  $\gamma$ , dan  $\text{degree}$  berperan penting dalam mengoptimalkan performa model.

## Saran

Penelitian selanjutnya disarankan untuk memfokuskan pemodelan pada variabel curah hujan sebagai variabel endogen tanpa melibatkan variabel eksogen seperti temperatur, kelembapan, dan kecepatan angin. Hal ini bertujuan untuk menyederhanakan model serta meningkatkan interpretabilitas hasil prediksi. Selain itu, jumlah data yang digunakan juga perlu diperhatikan. Penggunaan data selama satu tahun dapat lebih efektif karena data yang terlalu besar dengan banyak nilai nol atau ekstrem dapat meningkatkan error prediksi dan mengurangi akurasi model.

Evaluasi performa model juga sebaiknya lebih difokuskan pada metrik RMSE dan MAE, mengingat karakteristik data curah hujan yang sering mengandung nilai nol. Penggunaan metrik seperti MAPE kurang sesuai karena nilai nol dapat menghasilkan error yang tidak terdefinisi atau memberikan hasil yang kurang akurat. Dengan pendekatan ini, diharapkan penelitian selanjutnya dapat menghasilkan model yang lebih akurat, efisien, serta lebih mudah diinterpretasikan untuk keperluan prediksi curah hujan.

## DAFTAR PUSTAKA

- [1]. Aldrian, E., Pengkajian, B., Teknologi, P., & Thamrin, J. M. H. (n.d.). *Sistem Peringatan Dini Menghadapi Iklim Ekstrem Early Warning System for Climate Extreme*.
- [2]. Joewono, T. B., Susilo, Y. O., & Santosa, W. (2009). An Exploration of Public Transport Users' Attitudes and Preferences towards Various Policies in Indonesia: Some Preliminary Results. *Journal of the Eastern Asia Society for Transportation Studies*, 8. <https://doi.org/10.11175/easts.8.1230>.
- [3]. Hartatie, D., Taufika, R., & Achmad, P. B. (2021). Pengaruh Curah Hujan dan Pemupukan terhadap Produksi Tebu (*Saccharum officinarum L.*) di Pabrik Gula Asembagus Kabupaten Situbondo. *Jurnal Ilmiah Inovasi*, 21(2), 66–72. <https://doi.org/10.25047/jii.v21i2.2592>
- [4]. Juliati, Sulistiawaty., & Tiwow, V. A. (2023). Analisis Karakteristik Curah Hujan Dengan Menggunakan Klasifikasi Schmidt-Ferguson di Kota Makassar. *Jurnal Sains dan Pendidikan Fisika (JSPF)*, 19(2), 229-235. <https://ojs.unm.ac.id/jsdpf>

- [5]. Syamsiah, N. O., & Purwandani, I. (2021). Penerapan Ensemble Stacking untuk Peramalan Laba Bersih Bank Syariah Indonesia (BSI). *Building of Informatics, Technology and Science (BITS)*, 3(3), 295–301. <https://doi.org/10.47065/bits.v3i3.1017>
- [6]. Sugara , B., & Subekti, A. (2019). PENERAPAN SUPPORT VECTOR MACHINE (SVM) PADA SMALL DATASET UNTUK DETEksi DINI GANGGUAN AUTISME. *Jurnal Pilar Nusa Mandiri*, 15(2), 177–182. <https://doi.org/10.33480/pilar.v15i2.649>.
- [7]. Somantri, O., Wiro Sasmito, G., & Sobri Sungkar, M. (2014). Optimalisasi Neural Network dengan Bootstrap Aggregating (Bagging) untuk Penentuan Prediksi Harga Listrik. *Scientific Journal of Informatics*, 1(2), 2407–7658. <http://journal.unnes.ac.id/nju/index.php/sji>
- [8]. Dachi, J. M. A. S., & Sitompul, P. (2023). Analisis Perbandingan Algoritma XGBoost dan Algoritma Random Forest Ensemble Learning pada Klasifikasi Keputusan Kredit. *Jurnal Riset Rumpun Matematika dan Ilmu Pengetahuan Alam (JURRIMIPA)*, 2(2), 87–103. <https://doi.org/10.55606/jurrimipa.v2i2.1336>.
- [9]. Furi, R. P., Jondri, M. S., & Saepudin, D. (2015). Prediksi Financial Time Series Menggunakan Independent Component Analysis dan Support Vector Regression Studi Kasus : IHSG dan JII. 2(2) ISSN : 2355-9365
- [10]. Wijaya, A. B., Dewi, C., & Rahayudi, B. (2018). Peramalan Curah Hujan Menggunakan Metode High Order Fuzzy Time Series Multi Factors (Vol. 2, Issue 3). <http://j-ptiik.ub.ac.id>
- [11]. Kurniawati, A., Sabri Ahmad, M., Fhadli, M., & Lutfi, S. (2023). ANALISIS PERBANDINGAN METODE TIME SERIES FORECASTING UNTUK PREDIKSI PENJUALAN OBAT DI APOTEK (STUDI KASUS: KIMIA FARMA APOTEK TAKOMA) (Vol. 3, Issue 1).
- [12]. Lamabelawa, M. I. J. (2018). "PERBANDINGAN INTERPOLASI DAN EKSTRAPOLASI NEWTON UNTUK PREDIKSIDATA TIME SERIES". *JURNAL TEKNOLOGI INFORMASI*, 10(2) e-ISSN: 2620-7427.
- [13]. Sihombing, P. R., Suryadiningrat, S., Sunarjo, D. A., & Yuda, Y. P. A. C. (2023). Identifikasi Data Outlier (Penciran) dan Kenormalan Data Pada Data Univariat serta Alternatif Penyelesaiannya. *Jurnal Ekonomi Dan Statistik Indonesia*, 2(3), 307–316. <https://doi.org/10.11594/jesi.02.03.07>
- [14]. Whendasmoro, R. G., & Joseph, J. (2022). Analisis Penerapan Normalisasi Data Dengan Menggunakan Z-Score Pada Kinerja Algoritma K-NN. *JURIKOM (Jurnal Riset Komputer)*, 9(4), 872. <https://doi.org/10.30865/jurikom.v9i4.4526>
- [15]. Dini Maulana, N., Darma Setiawan, B., & Dewi, C. (2019). Implementasi Metode Support Vector Regression (SVR) Dalam Peramalan Penjualan Roti (Studi Kasus: Harum Bakery) (Vol. 3, Issue 3). <http://j-ptiik.ub.ac.id>
- [16]. Saragih, S. M., & Sembiring, P. (2022). ANALISIS PERBANDINGAN METODE ARIMA DAN DOUBLE EXPONENTIAL SMOOTHING DARI BROWN PADA PERAMALAN INFLASI DI INDONESIA. *Journal of Fundamental Mathematics and Applications (JFMA)*, 5(2), 176–191. <https://doi.org/10.14710/jfma.v5i2.15312>
- [17]. Prasetio, R. T., & Pratiwi, D. (2015). PENERAPAN TEKNIK BAGGING PADA ALGORITMA KLASIFIKASI UNTUK MENGATASI KETIDAKSEIMBANGAN KELAS DATASET MEDIS: Vol. II (Issue 2).
- [18]. Amansyah, I., Indra, J., Nurlaelasari, E., & Juwita, A. R. (2024). Prediksi Penjualan Kendaraan Menggunakan Regresi Linear: Studi Kasus pada Industri Otomotif di Indonesia. *INNOVATIVE: Journal Of Social Science Research*, 4(4), E-ISSN 2807-4238.