

Prediksi Curah Hujan Kota Bogor Menggunakan Random Forest, XGBoost, dan SVR Berdasarkan Parameter Meteorologi

Rainfall Prediction in Bogor City Using Random Forest, XGBoost, and SVR Based on Meteorological Parameters

Ika Wida Nuragustin^{*1}, Ersya Amelia², Brian Nugraha Wiyono³, Diah Septiani⁴

^{1,2,3,4} Program Studi S1 Sains Data, Universitas Telkom, Purwokerto, Jawa Tengah

e-mail: *ikawidanuragustin@gmail.com

Abstrak - Curah hujan merupakan salah satu parameter meteorologi yang memiliki pengaruh besar terhadap berbagai aktivitas dan berpotensi memicu bencana hidrometeorologi, khususnya di Kota Bogor yang dikenal memiliki intensitas curah hujan tinggi. Penelitian ini bertujuan untuk membandingkan performa algoritma Random Forest Regression, XGBoost Regression, dan Support Vector Regression (SVR) dalam memprediksi curah hujan serta menentukan model terbaik untuk melakukan forecasting. Data yang digunakan berupa data meteorologi yang terdiri atas temperatur, kelembapan, lama penyinaran matahari, dan kecepatan angin. Tahapan penelitian meliputi preprocessing data, feature engineering, pembagian data training dan testing, pelatihan model, serta evaluasi menggunakan metrik Mean Absolute Error (MAE), Root Mean Square Error (RMSE), dan koefisien determinasi (R^2). Hasil penelitian menunjukkan bahwa SVR memberikan performa terbaik dengan nilai MAE sebesar 2.7473, RMSE sebesar 4.8255, dan R^2 sebesar 0.9277, lebih baik dibandingkan Random Forest dan XGBoost. Model SVR kemudian digunakan untuk melakukan forecasting curah hujan 30 hari ke depan di Kota Bogor. Hasil forecasting menunjukkan bahwa curah hujan cenderung berada pada kategori tidak hujan hingga hujan sangat ringan, dengan beberapa peningkatan pada hari-hari tertentu. Temuan ini menunjukkan bahwa SVR mampu memodelkan hubungan non-linear pada data meteorologi secara efektif dan menghasilkan prediksi curah hujan yang akurat.

Kata kunci – Curah Hujan; Kota Bogor; Support Vector Regression; XGBoost, Random Forest

Abstract - Rainfall is a meteorological parameter that has a significant impact on various activities and has the potential to trigger hydrometeorological disasters, particularly in the city of Bogor, which is known for its high rainfall intensity. This study aims to compare the performance of the Random Forest Regression, XGBoost Regression, and Support Vector Regression (SVR) algorithms in predicting rainfall and to determine the best model for forecasting. The data used consists of meteorological data comprising temperature, humidity, duration of sunshine, and wind speed. The research stages include data preprocessing, feature engineering, splitting the data into training and testing sets, model training, and evaluation using the metrics Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and the coefficient of determination (R^2). The results of the study indicate that SVR delivers the best performance with an MAE value of 2.7473, an RMSE of 4.8255, and an R^2 of 0.9277, which is better than Random Forest and XGBoost. The SVR model was then used to forecast rainfall for the next 30 days in the city of Bogor. The forecast indicates that rainfall is likely to range from none to very light, with some increases on certain days. These findings demonstrate that SVR is capable of effectively modelling non-linear relationships in meteorological data and producing accurate rainfall predictions.

Keywords – Rainfall; Bogor City; Support Vector Regression; XGBoost, Random Forest

I. PENDAHULUAN

Indonesia terletak pada garis khatulistiwa dan termasuk wilayah beriklim tropis yang membentang antara 23,5° lintang utara hingga 23,5° lintang selatan. Kondisi geografis tersebut menyebabkan Indonesia memiliki suhu yang relatif stabil, namun dengan variasi curah hujan yang tinggi serta dipengaruhi oleh perubahan iklim global yang dapat mengubah pola hujan

di berbagai wilayah [1]. Salah satu wilayah dengan intensitas curah hujan tinggi adalah Kota Bogor di Provinsi Jawa Barat dengan luas wilayah 111,39 km². Kota Bogor dikenal sebagai “Kota Hujan” karena memiliki curah hujan yang sangat tinggi, bahkan dapat mencapai rata-rata 5.000 mm per tahun [2].

Tingginya curah hujan di Kota Bogor berkontribusi terhadap meningkatnya risiko bencana hidrometeorologi seperti banjir dan tanah longsor [3]. Dalam empat tahun terakhir, kejadian tanah longsor menyebabkan 10 korban jiwa, 1.385 terdampak, serta 170 rumah rusak, sedangkan banjir mengakibatkan 7 korban jiwa, 1.659 terdampak, dan 635 rumah terendam [4]. Kondisi tersebut menunjukkan bahwa curah hujan memiliki pengaruh signifikan terhadap tingkat kerentanan bencana, sehingga diperlukan sistem prediksi yang akurat untuk mendukung upaya mitigasi.

Meskipun demikian, prediksi curah hujan yang akurat masih menjadi tantangan yang kompleks karena melibatkan proses fisis atmosfer yang tidak linear serta dipengaruhi oleh berbagai variabel meteorologi [5]. Prediksi hujan umumnya dilakukan dalam skala waktu harian hingga mingguan [6]. Kompleksitas tersebut dipengaruhi oleh interaksi berbagai parameter meteorologi seperti suhu, kelembaban, tekanan udara, arah dan kecepatan angin, serta radiasi matahari [7].

Seiring perkembangan teknologi, pendekatan berbasis machine learning banyak digunakan untuk meningkatkan akurasi prediksi curah hujan dengan memanfaatkan data historis maupun data observasi secara langsung. Metode ini mampu menangkap pola hubungan non-linear antar variabel meteorologi yang sulit dimodelkan oleh pendekatan konvensional. Beberapa algoritma yang umum digunakan antara lain Random Forest, Support Vector Regression (SVR), dan XGBoost. Random Forest bekerja dengan menggabungkan beberapa decision tree untuk meningkatkan stabilitas prediksi serta mengurangi overfitting [8]. SVR menggunakan konsep hyperplane untuk memodelkan hubungan non-linear pada data berdimensi tinggi [9]. Sementara itu, XGBoost merupakan metode ensemble learning yang dikenal memiliki efisiensi tinggi serta kemampuan menangkap pola kompleks pada data [10].

Penelitian sebelumnya oleh Syahreza et al. [11] membandingkan Random Forest, Support Vector Regression (SVR), dan XGBoost dalam prediksi cuaca menggunakan data meteorologi dan data IoT. Hasil penelitian menunjukkan bahwa XGBoost memberikan performa terbaik dengan nilai MAE sebesar 0,3744, MSE sebesar 0,2278, dan R² sebesar 0,8183, diikuti oleh Random Forest dan SVR yang memiliki performa sedikit lebih rendah. Temuan tersebut menunjukkan bahwa ketiga algoritma mampu menangani pola data cuaca yang kompleks dengan baik, dengan pendekatan ensemble learning seperti XGBoost memiliki keunggulan dalam meningkatkan akurasi prediksi.

Meskipun penelitian tersebut telah menggunakan tiga algoritma yang sama, pendekatan yang dilakukan masih bersifat umum dan belum difokuskan pada wilayah dengan karakteristik curah hujan tinggi. Variasi kondisi geografis dan iklim suatu wilayah dapat memengaruhi pola distribusi data meteorologi serta kinerja model prediksi secara signifikan. Oleh karena itu, diperlukan kajian lebih lanjut pada wilayah dengan intensitas curah hujan tinggi untuk mengevaluasi stabilitas dan akurasi masing-masing algoritma secara lebih spesifik.

Penelitian lain oleh Syaharani et al. [12] menerapkan Random Forest untuk prediksi curah hujan di Kota Bogor dengan optimasi parameter menggunakan Grid Search dan validasi K-Fold Cross Validation. Hasil penelitian menunjukkan bahwa model Random Forest mencapai performa terbaik dengan RMSE sebesar 21,92 mm dan MAE sebesar 15,21 mm. Namun, penelitian tersebut masih terbatas pada penggunaan satu algoritma sehingga belum memberikan perbandingan dengan metode pembelajaran mesin lainnya.

Berdasarkan tinjauan tersebut, terdapat kesenjangan penelitian, yaitu belum adanya studi yang secara komprehensif membandingkan kinerja Random Forest, Support Vector Regression, dan XGBoost dalam satu penelitian untuk prediksi curah hujan pada data dengan karakteristik meteorologi yang kompleks di wilayah dengan intensitas hujan tinggi. Perbedaan karakteristik data dan kondisi geografis dapat memengaruhi performa model secara signifikan, sehingga diperlukan evaluasi multi-algoritma untuk memperoleh hasil yang lebih optimal.

Oleh karena itu, penelitian ini mengusulkan pendekatan komparatif dengan menerapkan Random Forest, Support Vector Regression, dan XGBoost dalam satu kerangka evaluasi untuk prediksi curah hujan di Kota Bogor berdasarkan parameter meteorologi seperti temperatur, kelembapan, lama penyinaran matahari, dan kecepatan angin. Pendekatan ini memungkinkan analisis performa masing-masing model secara lebih objektif dalam menangani data meteorologi yang kompleks. Hasil penelitian ini diharapkan dapat memberikan kontribusi dalam pemilihan model prediksi yang paling akurat serta mendukung upaya mitigasi bencana hidrometeorologi di Kota Bogor.

II. TINJAUAN TEORI

2.1 Curah Hujan

Curah hujan adalah jumlah air hujan yang jatuh pada permukaan bumi dalam suatu wilayah dan periode waktu tertentu. Curah hujan diukur sebagai ketinggian air hujan yang tertampung dalam alat penakar pada permukaan datar dengan satuan milimeter (mm), di mana 1 mm menunjukkan tinggi air setara 1 liter per meter persegi [13]. Variasi curah hujan antar wilayah dipengaruhi oleh kondisi geografis dan faktor meteorologi seperti suhu, tekanan udara, kelembapan, serta kecepatan dan arah angin. Tekanan udara berperan penting dalam distribusi curah hujan karena perubahan tekanan dapat memengaruhi pola angin yang berdampak pada pergerakan massa udara dan pembentukan awan hujan. Selain itu, ketinggian wilayah juga memengaruhi kondisi atmosfer yang berhubungan dengan perbedaan curah hujan antar daerah [14].

2.2 Algoritma Random Forest

Random Forest adalah algoritma *ensemble learning* yang terdiri dari kumpulan *decision tree* yang dibangun menggunakan *bootstrap sampling*. Hasil prediksi diperoleh dari rata-rata seluruh pohon pada kasus regresi. Adapun persamaan rumus Random Forest ditunjukkan pada Persamaan (1).

$$\hat{y} = \frac{1}{N} \sum_{i=1}^N f_i(x) \quad (1)$$

Metode ini mampu mengurangi *overfitting* dan meningkatkan akurasi prediksi pada dataset kompleks. Penelitian oleh Syaharani et al. [12] menunjukkan bahwa Random Forest efektif dalam prediksi curah hujan dengan error yang relatif rendah.

2.3 Algoritma XGBoost

XGBoost (*Extreme Gradient Boosting*) merupakan algoritma berbasis *ensemble learning* yang menggunakan pendekatan *gradient boosting* untuk membangun model secara bertahap. Setiap model baru dibentuk untuk memperbaiki kesalahan dari model sebelumnya dengan meminimalkan fungsi loss secara iteratif. XGBoost dikenal memiliki efisiensi tinggi serta kemampuan dalam menangani data yang bersifat kompleks dan non-linear. Adapun persamaan rumus XGBoost ditunjukkan pada Persamaan (2).

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i) \quad (2)$$

Fungsi objektif:

$$Obj = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k) \quad (3)$$

dengan:

$l(y_i, \hat{y}_i)$ adalah loss function

$\Omega(f_k)$ adalah regularisasi untuk mengontrol kompleksitas model

Metode ini mampu meningkatkan akurasi prediksi melalui proses boosting yang berfokus pada perbaikan error secara bertahap. Penelitian oleh Sangaji dan Sutabri [15] menunjukkan bahwa XGBoost memberikan performa yang baik dalam prediksi curah hujan.

2.4 Algoritma Support Vector Machine

Support Vector Regression (SVR) merupakan pengembangan dari *Support Vector Machine* (SVM) yang digunakan untuk menyelesaikan permasalahan regresi dengan cara mencari *hyperplane* terbaik yang memiliki margin kesalahan minimum dalam batas toleransi tertentu (*epsilon-insensitive*). SVR mampu menangani hubungan non-linear antara variabel input dan output dengan menggunakan kernel trick. Adapun persamaan rumus SVR ditunjukkan pada Persamaan (4) dan (5).

$$f(x) = w^T x + b \quad (4)$$

$$|y_i - f(x_i)| \leq \varepsilon \quad (5)$$

Metode ini efektif dalam menangani data berdimensi tinggi dan pola non-linear, namun memiliki keterbatasan pada waktu komputasi untuk dataset berukuran besar. Penelitian oleh Siregar [16] menunjukkan bahwa SVR mampu memberikan hasil prediksi yang optimal.

2.5 Evaluasi Model

Pada penelitian ini digunakan tiga metrik evaluasi yaitu *Mean Absolute Error* (MAE), *Root Mean Squared Error* (RMSE), dan *Coefficient of Determination* (R^2). MAE digunakan untuk mengukur rata-rata kesalahan absolut antara nilai aktual dan prediksi [17]. Adapun persamaan rumus MAE ditunjukkan pada Persamaan (6).

$$MAE = \frac{1}{n} \sum |y_i - \hat{y}_i| \quad (6)$$

RMSE digunakan untuk mengukur besarnya error dengan memberikan penalti lebih besar pada kesalahan yang tinggi [17]. Adapun persamaan rumus RMSE ditunjukkan pada Persamaan (7).

$$RMSE = \sqrt{\frac{1}{n} \sum (y_i - \hat{y}_i)^2} \quad (7)$$

R^2 digunakan untuk mengukur kemampuan model dalam menjelaskan variasi data [17]. Adapun persamaan rumus R^2 ditunjukkan pada Persamaan (8).

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} \quad (8)$$

III. METODE

Penelitian ini melalui beberapa tahapan utama, yaitu *data collection*, *preprocessing*, *exploratory data analysis* (EDA), *feature engineering*, *data splitting*, *modeling*, *evaluation*, dan *forecasting*. Data yang digunakan diperoleh dari data *online* BMKG (<https://dataonline.bmkg.go.id>) melalui Stasiun Klimatologi Jawa Barat. Dataset mencakup data meteorologi harian dengan total 730 pengamatan pada periode 29 Mei 2024 hingga 28 Mei 2026. Variabel yang digunakan dalam penelitian ini disajikan pada Tabel 1.

Tabel 1. Variabel Penelitian

Notasi	Variabel	Tipe Data	Satuan
X1	Temperatur Rata-rata (Tavg)	Numerik	°C
X2	Kelembapan Rata-rata (RH_avg)	Numerik	%
X3	Lama Penyinaran Matahari (ss)	Numerik	jam
X4	Kecepatan Angin Rata-rata (ff_avg)	Numerik	m/s
X5	Tanggal	Date	-
Y	Curah Hujan (RR)	Numerik	mm

Data yang diperoleh kemudian melalui tahap *preprocessing*, pada tahap ini, kolom tanggal terlebih dahulu dikonversi ke format *datetime* dan data diurutkan berdasarkan waktu pengamatan untuk menjaga struktur deret waktu. Selanjutnya dilakukan pembersihan terhadap nilai tidak valid yang umum terdapat pada data BMKG, yaitu nilai 8888 yang merepresentasikan data tidak terukur dan nilai 9999 yang menunjukkan tidak dilakukannya pengukuran pada periode tertentu, sehingga seluruh nilai tersebut dikonversi menjadi *missing value* (NaN). Tahap berikutnya adalah penanganan *missing value*. Variabel numerik kemudian diolah dengan metode interpolasi linear untuk mempertahankan pola temporal data. Sisa nilai kosong yang masih terdapat pada awal atau akhir deret waktu diisi menggunakan nilai median masing-masing variabel. Setelah proses ini dilakukan, seluruh data dinyatakan lengkap tanpa *missing value* dan siap digunakan untuk tahap analisis selanjutnya.

Exploratory Data Analysis (EDA) dilakukan untuk memahami karakteristik data serta pola hubungan antar variabel. EDA mencakup analisis distribusi variabel dengan histogram, serta deteksi *outlier* menggunakan *boxplot*. Selain itu, dilakukan analisis distribusi dan tren curah hujan terhadap waktu, termasuk pola musiman melalui agregasi bulanan dan perbandingan antar tahun. Hubungan antar variabel dianalisis menggunakan korelasi dan visualisasi regresi untuk melihat pengaruh masing-masing fitur terhadap curah hujan. Hasil EDA digunakan sebagai dasar dalam proses *feature engineering* dan pemodelan.

Tahap *feature engineering* dilakukan untuk meningkatkan kemampuan model dalam menangkap pola dan karakteristik data curah hujan. Pada tahap ini, ditambahkan fitur waktu berupa bulan (*Bulan*) untuk merepresentasikan pola musiman. Selain itu, dibuat fitur *lag* yang terdiri dari curah hujan 1–3 hari sebelumnya (*Curah_Hujan_Lag1*, *Curah_Hujan_Lag2*, dan *Curah_Hujan_Lag3*) serta fitur *lag* untuk temperatur (*Temp_Lag1*) dan kelembapan (*Hum_Lag1*) guna memasukkan informasi historis ke dalam model. Untuk menangkap tren jangka pendek, ditambahkan fitur *rolling mean* curah hujan selama tiga hari terakhir (*Rain_Roll3*). Selanjutnya, dibuat fitur interaksi antarvariabel cuaca, yaitu interaksi antara temperatur dan kelembapan (*Temp_Hum*) serta kelembapan dan kecepatan angin (*Hum_Wind*), untuk memperkaya informasi hubungan antarvariabel. Setelah seluruh fitur dibentuk, dilakukan pemilihan variabel untuk pemodelan, di mana fitur (X) terdiri dari temperatur, kelembapan, lama penyinaran matahari, kecepatan angin, bulan, fitur *lag* curah hujan, fitur *lag* temperatur dan kelembapan, *rolling mean* curah hujan, serta fitur interaksi antarvariabel cuaca, sedangkan variabel target (y) yang digunakan adalah curah hujan.

Selanjutnya data dibagi menjadi data latih dan data uji berdasarkan urutan waktu dengan proporsi 80% untuk data latih dan 20% untuk data uji. Pembagian dilakukan tanpa pengacakan untuk menjaga karakteristik data deret waktu dan menghindari *data leakage*. Pada tahap ini juga dilakukan normalisasi data menggunakan *StandardScaler* khusus untuk model yang sensitif terhadap skala, yaitu *Support Vector Regression* (SVR). Penelitian ini menggunakan tiga algoritma *machine learning*, yaitu *Random Forest Regressor*, *Support Vector Regression* (SVR), dan *XGBoost Regressor*. Ketiga model dilatih menggunakan data latih dan diuji menggunakan data uji untuk memprediksi nilai curah hujan. Model SVR menggunakan data yang telah dinormalisasi, sedangkan Random Forest dan XGBoost menggunakan data asli tanpa *scaling*.

Setelah tahap pemodelan dilakukan, kinerja model dievaluasi menggunakan *Mean Absolute Error* (MAE), *Root Mean Squared Error* (RMSE), dan *R-squared* (R^2). Selain evaluasi numerik, dilakukan juga visualisasi perbandingan antara nilai aktual dan hasil prediksi untuk melihat kedekatan pola prediksi terhadap data sebenarnya. Berdasarkan hasil evaluasi tersebut, dipilih model terbaik yang memiliki performa paling optimal untuk digunakan pada tahap selanjutnya. Model terbaik yang diperoleh kemudian digunakan untuk melakukan *forecasting* atau peramalan nilai curah hujan untuk 30 hari ke depan menggunakan pendekatan *recursive forecasting*.

IV. HASIL DAN PEMBAHASAN

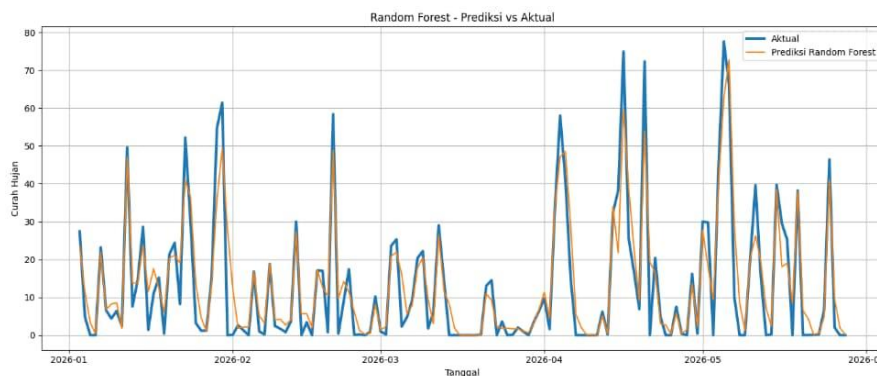
4.1 Random Forest Regression

Pada tahap ini dilakukan pemodelan menggunakan algoritma *Random Forest Regression*, model Random Forest dibangun menggunakan parameter sebanyak 500 pohon keputusan dengan kedalaman maksimum pohon sebesar 8 untuk mengontrol kompleksitas model dan mengurangi risiko *overfitting*. Selain itu, digunakan *random state* sebesar 42 untuk memastikan hasil pelatihan model bersifat konsisten dan dapat direproduksi pada setiap percobaan. Model kemudian dilatih menggunakan data *training* dan diuji pada data *testing* untuk mengevaluasi kemampuan generalisasi model terhadap data yang belum pernah dilihat sebelumnya. Hasil evaluasi Random Forest disajikan pada Tabel 2 berikut.

Tabel 2. Hasil Evaluasi Model Random Forest

Metode	MAE	RMSE	R ²
Random Forest	4.4759	6.7308	0.8593

Berdasarkan Tabel 2, nilai MAE sebesar 4.4759 menunjukkan bahwa rata-rata selisih absolut antara nilai curah hujan aktual dan hasil prediksi model Random Forest berada pada kisaran 4.48 satuan, yang mencerminkan tingkat kesalahan prediksi tanpa memperhatikan arah kesalahan. Nilai RMSE sebesar 6.7308 mengindikasikan besarnya kesalahan prediksi dengan memberikan penalti lebih besar pada error yang tinggi, sehingga menunjukkan adanya beberapa prediksi dengan penyimpangan yang cukup besar meskipun sebagian besar error masih relatif kecil. Sementara itu, nilai R² sebesar 0.8593 menunjukkan bahwa model mampu menjelaskan sekitar 85.93% variasi data curah hujan, yang berarti sebagian besar pola dalam data berhasil ditangkap oleh model, namun masih terdapat sekitar 14.07% variasi yang belum dapat dijelaskan oleh variabel yang digunakan. Selanjutnya dilakukan visualisasi untuk melihat kesesuaian antara nilai aktual dan hasil prediksi model. Hasil visualisasi ditunjukkan pada Gambar 1.

**Gambar 1.** Perbandingan Nilai Aktual dan Prediksi Random Forest

Berdasarkan Gambar 1, nilai prediksi Random Forest mampu mengikuti pola data aktual dengan baik, ditunjukkan oleh kedekatan antara garis prediksi dan garis aktual pada sebagian besar periode pengamatan. Meskipun masih terdapat sedikit perbedaan pada beberapa puncak curah hujan, model tetap dapat menangkap tren dan fluktuasi data dengan cukup akurat.

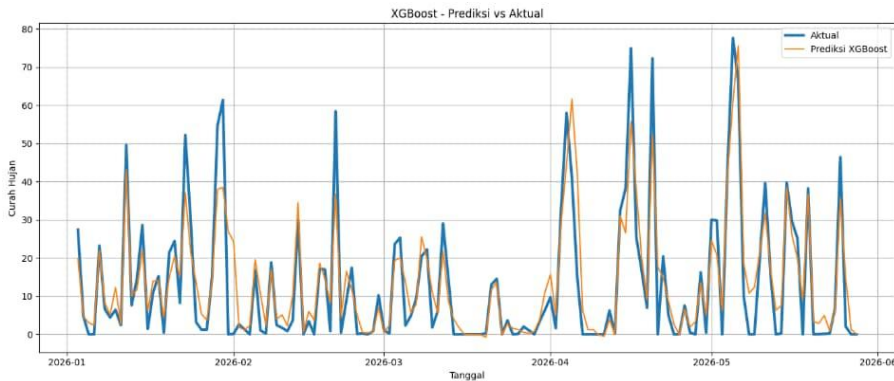
4.2 XGBoost Regression

Model XGBoost dibangun dengan jumlah *boosting tree* sebanyak 500 untuk meningkatkan kemampuan model dalam mempelajari pola data yang kompleks. *Learning rate* sebesar 0.03 digunakan untuk mengontrol kecepatan pembelajaran agar model tidak terlalu agresif dalam menyesuaikan data. Kedalaman maksimum pohon ditetapkan sebesar 5 untuk menjaga kompleksitas model agar tidak terjadi *overfitting*. Selain itu, digunakan subsampling sebesar 0.8 dan *colsample_bytree* sebesar 0.8 untuk meningkatkan generalisasi model dengan cara hanya menggunakan sebagian data dan fitur pada setiap iterasi. Fungsi objektif yang digunakan adalah regresi dengan error kuadrat (*reg:squarederror*), serta *random state* sebesar 42 untuk memastikan hasil yang konsisten dan dapat direproduksi. Model kemudian dilatih menggunakan data training dan diuji menggunakan data testing untuk mengevaluasi kemampuan generalisasi model terhadap data yang belum pernah dilihat sebelumnya. Hasil evaluasi model XGBoost disajikan pada Tabel 3 berikut:

Tabel 3. Hasil Evaluasi Model XGBoost

Metode	MAE	RMSE	R ²
XGBoost	5.1195	7.6989	0.8159

Berdasarkan Tabel 3, nilai MAE sebesar 5.1195 menggambarkan rata-rata selisih absolut antara hasil prediksi dan nilai aktual curah hujan yang berada pada kisaran 5.12 satuan, sehingga menunjukkan besarnya deviasi prediksi tanpa memperhitungkan arah kesalahan. Kemudian, RMSE yang bernilai 7.6989 memperlihatkan bahwa kesalahan prediksi memiliki variasi yang cukup lebar karena metrik ini lebih sensitif terhadap error yang besar, sehingga mengindikasikan adanya beberapa prediksi yang meleset cukup jauh dari nilai sebenarnya. Di sisi lain, nilai R² sebesar 0.8159 mengisyaratkan bahwa sekitar 81.59% variasi data curah hujan dapat dijelaskan oleh model, yang berarti pola utama dalam data berhasil dipelajari dengan cukup baik meskipun masih terdapat bagian variasi yang belum terakomodasi oleh model. Selanjutnya dilakukan visualisasi untuk melihat kesesuaian antara nilai aktual dan hasil prediksi model. Hasil visualisasi ditunjukkan pada Gambar 2.



Gambar 2. Perbandingan Nilai Aktual dan Prediksi XGBoost

Berdasarkan Gambar 2, nilai prediksi XGBoost mampu mengikuti pola perubahan curah hujan yang terjadi pada data aktual. Sebagian besar puncak dan penurunan curah hujan berhasil diprediksi dengan cukup baik, meskipun terdapat beberapa periode di mana nilai prediksi terlihat lebih tinggi atau lebih rendah dibandingkan nilai aktual. Perbedaan tersebut menunjukkan bahwa model masih mengalami kesalahan pada beberapa titik pengamatan, terutama saat terjadi fluktuasi curah hujan yang cukup tajam.

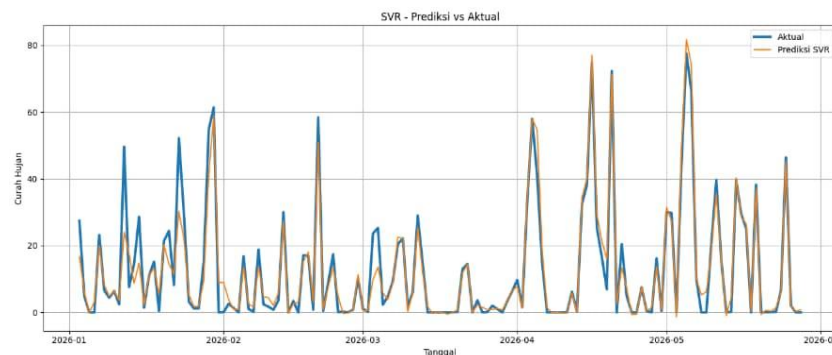
4.3 Support Vector Regression

Pada pemodelan *Support Vector Regression* (SVR) digunakan kernel radial basis *function* (RBF) untuk menangani pola data yang tidak linier. Parameter regularisasi ditetapkan sebesar 100 untuk mengontrol keseimbangan antara kesalahan pelatihan dan kompleksitas model. Parameter gamma menggunakan nilai otomatis (*scale*) untuk menyesuaikan pengaruh setiap data terhadap model. Selain itu, digunakan epsilon sebesar 0.1 sebagai batas toleransi kesalahan prediksi. Karena SVR sensitif terhadap skala data, maka data yang digunakan dalam proses pelatihan adalah data yang telah dilakukan normalisasi (*scaling*). Model kemudian dilatih menggunakan data *training* dan diuji menggunakan data *testing* untuk mengevaluasi kemampuan generalisasi model terhadap data yang belum pernah dilihat sebelumnya. Hasil evaluasi model SVR disajikan pada Tabel 4 berikut:

Tabel 4. Hasil Evaluasi Model SVR

Metode	MAE	RMSE	R ²
SVR	2.7473	4.8255	0.9277

Berdasarkan Tabel 4, MAE sebesar 2.7473 menunjukkan bahwa rata-rata selisih absolut antara nilai aktual dan hasil prediksi SVR berada pada kisaran 2.75 satuan, yang mencerminkan tingkat deviasi prediksi yang relatif kecil dalam menangkap pola data. Selanjutnya, RMSE yang bernilai 4.8255 menggambarkan besarnya kesalahan prediksi dengan sensitivitas yang lebih tinggi terhadap error besar, sehingga menunjukkan bahwa penyimpangan prediksi masih ada namun tidak terlalu ekstrem secara umum. Sementara itu, nilai R² sebesar 0.9277 mengindikasikan bahwa model mampu menjelaskan sekitar 92.77% variasi data curah hujan, yang berarti sebagian besar pola utama dalam data berhasil direpresentasikan dengan baik oleh model SVR. Untuk memvalidasi hasil evaluasi numerik, dilakukan visualisasi perbandingan antara nilai aktual dan nilai prediksi model SVR. Hasil visualisasi ditunjukkan pada Gambar 3.

**Gambar 3.** Perbandingan Nilai Aktual dan Prediksi SVR

Berdasarkan Gambar 3, nilai prediksi SVR terlihat sangat dekat dengan nilai aktual pada sebagian besar periode pengamatan. Garis prediksi mampu mengikuti pola kenaikan dan penurunan curah hujan dengan baik, termasuk pada beberapa titik dengan nilai curah hujan yang tinggi. Kedekatan antara kedua garis menunjukkan bahwa model SVR memiliki kemampuan yang baik dalam menangkap pola data curah hujan dan menghasilkan prediksi yang relatif akurat.

4.4 Perbandingan Performa Model

Untuk mengevaluasi kinerja model secara komprehensif, dilakukan perbandingan terhadap tiga algoritma yang digunakan dalam penelitian ini, yaitu Random Forest, XGBoost, dan Support Vector Regression (SVR). Evaluasi ini bertujuan untuk melihat sejauh mana masing-masing model mampu memberikan hasil prediksi yang akurat berdasarkan metrik kesalahan (MAE dan RMSE) serta kemampuan model dalam menjelaskan variasi data (R²). Hasil pengukuran dari ketiga model tersebut kemudian dirangkum dalam Tabel 5 sebagai dasar untuk analisis perbandingan performa.

Tabel 5. Perbandingan Performa Model

Model	Metrix			Best Model
	MAE	RMSE	R ²	
Random Forest	4.4759	6.7308	0.8593	SVR
XGBoost	5.1195	7.6989	0.8159	SVR
SVR	2.7473	4.8255	0.9277	SVR

Berdasarkan Tabel 5, terlihat perbedaan karakteristik performa yang cukup jelas antar model melalui tiga metrik evaluasi utama. Random Forest menunjukkan performa yang relatif stabil dengan MAE 4.4759 dan RMSE 6.7308, yang mengindikasikan kesalahan prediksi masih berada pada tingkat sedang, sementara R^2 sebesar 0.8593 menandakan bahwa model sudah mampu menangkap pola utama data, namun belum sepenuhnya presisi pada variasi yang lebih kompleks. XGBoost memperlihatkan penurunan performa dengan MAE 5.1195 dan RMSE 7.6989, yang menunjukkan bahwa model ini lebih sering menghasilkan deviasi yang lebih besar dalam prediksi, dan hal ini tercermin pula pada R^2 sebesar 0.8159 yang lebih rendah, sehingga kemampuan dalam merepresentasikan variasi data menjadi paling terbatas di antara ketiga model.

Berbeda dengan kedua model tersebut, SVR memberikan hasil yang paling konsisten di seluruh metrik, dengan MAE 2.7473 dan RMSE 4.8255 yang menunjukkan tingkat kesalahan prediksi yang jauh lebih rendah, baik pada kesalahan rata-rata maupun kesalahan yang bersifat ekstrem. Nilai R^2 sebesar 0.9277 memperkuat indikasi bahwa SVR mampu menangkap pola hubungan dalam data curah hujan secara lebih menyeluruh, termasuk variasi yang lebih halus. Kombinasi rendahnya error dan tingginya kemampuan penjelasan varians menunjukkan bahwa SVR lebih adaptif terhadap pola non-linear pada data cuaca, sehingga menghasilkan prediksi yang lebih akurat dan stabil dibandingkan dua model lainnya.

4.5 Forecasting Curah Hujan

Berdasarkan hasil analisis sebelumnya, model SVR menunjukkan performa terbaik dibandingkan model lainnya, baik dari sisi error maupun kemampuan menjelaskan variasi data. Dengan pertimbangan tersebut, tahap selanjutnya dilakukan forecasting curah hujan 30 hari ke depan menggunakan model SVR sebagai model yang paling representatif untuk data curah hujan Kota Bogor. Hasil Forecasting ditunjukkan pada Tabel 6 berikut.

Tabel 6. Forecasting Curah Hujan 30 Hari Kedepan di Kota Bogor Menggunakan SVR

No	Tanggal	Prediksi Curah Hujan (mm)	No	Tanggal	Prediksi Curah Hujan (mm)
1	2026-05-29	1.5378	16	2026-06-13	0.0000
2	2026-05-30	0.0000	17	2026-06-14	0.2071
3	2026-05-31	0.2587	18	2026-06-15	0.3403
4	2026-06-01	1.0769	19	2026-06-16	1.3068
5	2026-06-02	0.0000	20	2026-06-17	0.0909
6	2026-06-03	0.2717	21	2026-06-18	0.2945
7	2026-06-04	0.2724	22	2026-06-19	0.8504
8	2026-06-05	0.2798	23	2026-06-20	0.0984
9	2026-06-06	0.1863	24	2026-06-21	0.9332
10	2026-06-07	0.3543	25	2026-06-22	0.6034
11	2026-06-08	0.1623	26	2026-06-23	0.0000
12	2026-06-09	0.1695	27	2026-06-24	0.6170
13	2026-06-10	0.3299	28	2026-06-25	0.4237
14	2026-06-11	0.1035	29	2026-06-26	0.1109
15	2026-06-12	0.2639	30	2026-06-27	0.0220

Berdasarkan Tabel 6, hasil forecasting curah hujan 30 hari ke depan di Kota Bogor menunjukkan bahwa nilai prediksi didominasi oleh angka yang sangat rendah, berkisar

antara 0 mm hingga 1,5378 mm per hari. Nilai prediksi sebesar 0 mm mengindikasikan bahwa pada hari tersebut diperkirakan tidak terjadi hujan, sedangkan nilai prediksi antara 0,1–5 mm/hari termasuk kategori hujan sangat ringan. Beberapa hari dengan prediksi tertinggi, seperti 29 Mei 2026 (1,5378 mm), 1 Juni 2026 (1,0769 mm), 16 Juni 2026 (1,3068 mm), 19 Juni 2026 (0,8504 mm), 21 Juni 2026 (0,9332 mm), dan 24 Juni 2026 (0,6170 mm), masih berada dalam kategori hujan sangat ringan. Hal ini menunjukkan bahwa selama periode 30 hari ke depan tidak diperkirakan terjadi hujan dengan intensitas ringan, sedang, maupun lebat. Dengan demikian, kondisi curah hujan di Kota Bogor diprediksi relatif stabil dan didominasi oleh kondisi tidak hujan hingga hujan sangat ringan, sehingga potensi terjadinya hujan dengan intensitas tinggi pada periode prediksi tergolong rendah.

Informasi yang dihasilkan dari *forecasting* curah hujan dapat dimanfaatkan sebagai pendukung pengambilan keputusan yang berkaitan dengan kondisi cuaca di Kota Bogor. Meskipun hasil prediksi menunjukkan curah hujan yang relatif rendah selama 30 hari ke depan, informasi mengenai waktu terjadinya peningkatan curah hujan tetap penting untuk mengidentifikasi perubahan pola cuaca pada hari-hari tertentu. Mengingat Kota Bogor merupakan wilayah dengan intensitas curah hujan yang tinggi dan memiliki kerentanan terhadap bencana hidrometeorologi, hasil prediksi ini dapat dimanfaatkan untuk meningkatkan kesiapsiagaan terhadap potensi bencana banjir dan tanah longsor yang dipicu oleh peningkatan curah hujan.

Selain itu, hasil *forecasting* juga dapat dimanfaatkan pada berbagai sektor yang sensitif terhadap kondisi cuaca. Contohnya pada sektor pertanian, informasi curah hujan dapat digunakan untuk menentukan waktu tanam yang lebih tepat, mengatur kebutuhan irigasi, serta merencanakan jadwal pemupukan dan panen agar lebih sesuai dengan kondisi cuaca yang diperkirakan terjadi. Pada sektor transportasi, prediksi curah hujan dapat membantu dalam perencanaan perjalanan dan aktivitas operasional yang dipengaruhi oleh kondisi cuaca. Sementara itu, pada sektor pariwisata, informasi prediksi cuaca dapat menjadi pertimbangan dalam penyusunan jadwal kegiatan wisata, terutama pada destinasi wisata alam yang banyak terdapat di wilayah Bogor. Hasil *forecasting* tidak hanya memberikan gambaran kondisi curah hujan di Kota Bogor pada periode mendatang, tetapi juga menghasilkan informasi yang dapat dimanfaatkan sebagai dasar pengambilan keputusan yang lebih efektif dan berbasis data.

V. KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan, algoritma *Support Vector Regression* (SVR) merupakan model terbaik dalam prediksi curah hujan di Kota Bogor berdasarkan parameter meteorologi yang digunakan. Model SVR menghasilkan nilai MAE sebesar 2.7473, RMSE sebesar 4.8255, dan R^2 sebesar 0.9277, yang menunjukkan kemampuan prediksi yang baik dengan tingkat kesalahan yang rendah serta kemampuan yang tinggi dalam menjelaskan variasi data curah hujan. Hasil *forecasting* menggunakan model SVR menunjukkan bahwa curah hujan di Kota Bogor selama 30 hari ke depan didominasi oleh kondisi tidak hujan hingga hujan sangat ringan, dengan nilai prediksi berkisar antara 0 mm hingga 1,5378 mm per hari dan fluktuasi yang relatif kecil. Temuan ini menunjukkan bahwa SVR efektif dalam menangkap pola non-linear pada data meteorologi sehingga dapat digunakan sebagai model yang andal untuk prediksi curah hujan di Kota Bogor.

5.2 Saran

Penelitian selanjutnya disarankan menggunakan data dengan periode yang lebih panjang agar pola curah hujan dapat dipelajari secara lebih komprehensif. Selain itu, penambahan variabel meteorologi maupun faktor iklim lainnya serta optimasi parameter model dapat dilakukan untuk meningkatkan akurasi prediksi. Pengujian menggunakan metode *machine learning* dan *deep learning* lainnya juga dapat dilakukan sebagai pembandingan untuk memperoleh model prediksi curah hujan yang lebih optimal.

VI. DAFTAR PUSTAKA

- [1] R. C. Tarumingkeng, *Pengaruh Perubahan Iklim*, Bogor, Indonesia: RUDYCT e-PRESS, 2024.
- [2] M. Mahfudz, B. Riadi, dan I. Rifaldi, "Pemetaan Area Potensi Banjir Berdasarkan Topographic Wetness Index (TWI) di Kecamatan Cigudeg Kabupaten Bogor," *Geomatika*, vol. 28, no. 1, pp. 13–20, 2022.
- [3] M. Azizah, A. Subiyanto, S. Triutomo, dan D. Wahyuni, "Pengaruh Perubahan Iklim Terhadap Bencana Hidrometeorologi di Kecamatan Cisarua - Kabupaten Bogor," *PENDIPA Journal of Science Education*, vol. 6, no. 2, pp. 541–546, 2022.
- [4] BNPB, *Data Informasi Bencana Indonesia*, 2024. [Online]. Available: <https://dibi.bnpb.go.id/>
- [5] R. Al Fauzi, "Analisis tingkat kerawanan banjir Kota Bogor menggunakan metode overlay dan scoring berbasis sistem informasi geografis," *Geomedia*, vol. 20, no. 2, pp. 96–107, 2022.
- [6] M. Yusuf, A. Setyanto, dan K. Aryasa, "Analisis Prediksi Curah Hujan Bulanan Wilayah Kota Sorong Menggunakan Metode Multiple Regression," *Jurnal Sains Komputer dan Informatika*, vol. 6, no. 1, pp. 405–417, 2022.
- [7] M. A. Hasanah, S. Soim, A. S. Handayani, et al., "Implementasi CRISP-DM Model Menggunakan Metode Decision Tree dengan Algoritma CART untuk Prediksi Curah Hujan Berpotensi Banjir," *Journal of Applied Informatics and Computing*, vol. 5, no. 2, pp. 103–108, 2021.
- [8] A. M. Priyatno et al., "Comparison Random Forest Regression and Linear Regression for Forecasting BBKA Stock Price," *Jurnal Teknik Industri Terintegrasi*, vol. 6, no. 3, pp. 718–732, 2023.
- [9] A. Setiawan dan R. R. Suryono, "Analisis Sentimen Ibu Kota Nusantara Menggunakan Algoritma Support Vector Machine dan Naïve Bayes," *Edumatic: Jurnal Pendidikan Informatika*, vol. 8, no. 1, pp. 183–192, 2024.
- [10] D. N. Muhammadiyah et al., "Students Final Academic Score Prediction Using Boosting Regression Algorithms," *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika (JITEKI)*, vol. 10, no. 1, pp. 154–165, 2024.
- [11] A. Syahreza, N. K. Ningrum, dan M. A. Syahrazy, "Perbandingan Kinerja Model Prediksi Cuaca: Random Forest, Support Vector Regression, dan XGBoost," *Edumatic: Jurnal Pendidikan Informatika*, vol. 8, no. 2, pp. 526–534, 2024.
- [12] M. A. Syaharani, B. A. Dermawan, dan R. I. Adam, "Prediksi Curah Hujan Kota Bogor Menggunakan Algoritma Random Forest," *Jurnal Sistem Informasi dan Aplikasi*, vol. 3, no. 1, pp. 52–62, 2025.
- [13] G. A. Windari dan S. Sudarti, "Mekanisme terjadinya hujan dan pengaruhnya terhadap lingkungan," *Jurnal Teknologi Lingkungan UNMUL*, vol. 8, no. 2, pp. 11–20, 2024.
- [14] K. H. A. Kuncoro et al., "Analisis Pengaruh Curah Hujan terhadap Rata-Rata Suhu, Kecepatan Angin, dan Kelembapan di Kota Bandar Lampung," *Jurnal Ilmiah Teknosains*, vol. 11, no. 1, 2025.
- [15] D. Sangaji dan T. Sutabri, "Analisis XGBoost dan Random Forest untuk Prediksi Curah Hujan dalam Mendukung Mitigasi Karhutla," *Jurnal Pustaka AI*, vol. 5, no. 1, pp. 13–18, 2025.
- [16] N. A. Siregar, "Peramalan Curah Hujan di Kota Medan menggunakan Metode Support Vector Regression," *Journal of Informatics and Data Science (J-IDS)*, vol. 1, no. 1, pp. 1–5, 2022.
- [17] M. A. Dinata, H. Yenni, dan W. Agustin, "Prediksi Jumlah Titik Ruang Terbuka Hijau (RTH) Menggunakan Metode Regresi Linier dan Random Forest," *BETRIK*, vol. 16, no. 2, pp. 144–156, 2025.