

Prediksi Curah Hujan Menggunakan Random Forest dan VAR di Kediri Raya

Intan Luthfiyah Rahmah^a, Aryo Nugroho^b, Raja Yosia Manahan Trinitas Siregar^c

^aTeknik Informatika, Fakultas Ilmu Komputer, Universitas Narotama, rahmahfiyah@gmail.com

^bSistem Informasi, Fakultas Ilmu Komputer, Universitas Narotama, aryonugroho@narotama.ac.id

^cSistem Informasi, Fakultas Ilmu Komputer, Universitas Narotama, yosia19@mhs.fasilkom.narotama.ac.id

Submitted: 11-07-2025, Reviewed: 02-09-2025, Accepted 24-09-2025
<https://doi.org/10.47233/jteksis.v7i4.2141>

Abstract

Changes in rainfall patterns in the Kediri Raya region have affected the agricultural sector, infrastructure, and disaster mitigation. The presence of Dhoho Airport as a new economic hub increases the demand for precise weather prediction to ensure operational safety and risk management. This study proposes a daily rainfall prediction model using Random Forest (RF) and Vector Autoregression (VAR) with climatological data from BMKG, including average temperature, humidity, sunshine duration, and wind speed. The data were processed using logarithmic transformation to stabilize variance, followed by modeling and evaluation with the Mean Squared Error (MSE) metric. The results show that VAR is more effective in capturing temporal dynamics, while RF produces more stable predictions. These findings highlight that combining the two methods can provide more reliable results compared to a single approach. The developed model has the potential to be integrated into a local weather prediction information system to support hydrometeorological disaster mitigation, water resource management, and operational strategies at Dhoho Airport in Kediri.

Keywords: Rainfall prediction, Random Forest, Vector Autoregression, Dhoho Airport, Information system

Abstrak

Perubahan pola curah hujan di Kediri Raya berdampak pada sektor pertanian, infrastruktur, dan mitigasi bencana. Kehadiran Bandara Dhoho sebagai pusat ekonomi baru menambah kebutuhan prediksi cuaca yang presisi untuk mendukung keselamatan operasional dan perencanaan risiko. Penelitian ini mengusulkan model prediksi curah hujan harian menggunakan Random Forest (RF) dan Vector Autoregression (VAR) dengan data klimatologi dari BMKG, meliputi suhu rata-rata, kelembapan, durasi penyinaran matahari, dan kecepatan angin. Data diolah melalui transformasi logaritmik untuk menstabilkan variansi, kemudian dilakukan pemodelan dan evaluasi menggunakan Mean Squared Error (MSE). Hasil penelitian menunjukkan bahwa VAR lebih efektif menangkap dinamika temporal, sementara RF menghasilkan prediksi yang lebih stabil. Temuan ini menegaskan bahwa kombinasi kedua metode dapat memberikan hasil yang lebih reliabel dibandingkan pendekatan tunggal. Model yang dikembangkan berpotensi diintegrasikan dalam sistem informasi prediksi cuaca lokal untuk mendukung mitigasi bencana hidrometeorologi, pengelolaan sumber daya air, serta strategi operasional di Bandara Dhoho Kediri.

Keywords: Prediksi curah hujan, Random Forest, Vector Autoregression, Bandara Dhoho, Sistem Informasi

This work is licensed under Creative Commons Attribution License 4.0 CC-BY International license



PENDAHULUAN

Perubahan iklim yang semakin tidak menentu telah membuat pola curah hujan di wilayah tropis, termasuk Kediri Raya, sulit diprediksi. Kondisi ini meningkatkan risiko bencana hidrometeorologi seperti banjir, tanah longsor, dan gangguan pada sektor pertanian serta infrastruktur [1]. Salah satu infrastruktur vital yang terdampak adalah Bandara Dhoho Kediri, yang berfungsi sebagai simpul transportasi udara dan penunjang pertumbuhan ekonomi lokal. Dalam konteks tersebut, ketersediaan sistem prediksi curah hujan harian yang akurat menjadi sangat penting untuk mendukung keselamatan operasional bandara serta mitigasi risiko bencana [2].

Vector Autoregression (VAR) sering digunakan untuk menganalisis dinamika temporal

data klimatologi deret waktu, seperti suhu, kelembapan, penyinaran matahari, dan kecepatan angin [3]. Namun, metode ini kurang optimal dalam menangani hubungan non-linear yang umum terjadi pada sistem cuaca tropis [4]. Sebagai alternatif, Random Forest (RF) menawarkan fleksibilitas dalam mengelola hubungan non-linear dan telah banyak diaplikasikan pada studi prediksi iklim [5].

Integrasi VAR dan RF dalam model hybrid terbukti mampu meningkatkan akurasi prediksi curah hujan dengan menggabungkan keunggulan analisis statistik dan machine learning [6][7]. Pendekatan hybrid ini juga lebih adaptif terhadap variasi data dan memberikan keandalan lebih tinggi dibandingkan metode konvensional [8]. Seiring dengan pengoperasian Bandara Dhoho sebagai pusat

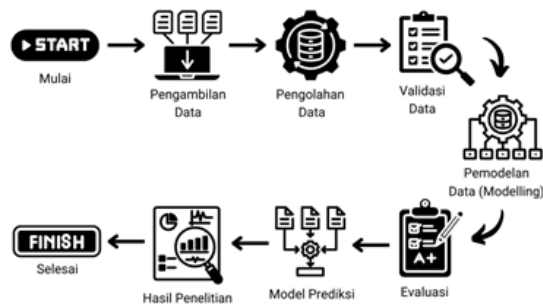
ekonomi baru, kebutuhan prediksi cuaca yang presisi semakin meningkat[9][10].

Penelitian sebelumnya menunjukkan keberhasilan RF dalam memprediksi curah hujan harian dengan akurasi tinggi di Jawa Timur[11], sementara VAR efektif memperkirakan dinamika cuaca jangka pendek di wilayah tropis[12]. Ombinasi kedua metode ini juga telah digunakan dalam sistem peringatan dini cuaca ekstrem yang adaptif terhadap perubahan iklim lokal[13]. Dengan dukungan data historis BMKG dan pemrosesan berbasis machine learning, model prediksi dapat dikembangkan secara kontekstual untuk mendukung mitigasi banjir dan manajemen infrastruktur transportasi udara[14][15].

Berdasarkan latar belakang tersebut, penelitian ini bertujuan mengembangkan model prediksi curah hujan harian di Kediri Raya menggunakan pendekatan gabungan VAR dan RF, dengan harapan dapat mendukung sistem informasi iklim lokal serta strategi mitigasi risiko bencana secara lebih presisi dan aplikatif.

METODE PENELITIAN

Dalam penelitian ini, metode yang digunakan mengikuti alur yang digambarkan pada Gambar 1.



Gambar 1. Metode Penelitian

Data yang digunakan dalam penelitian ini adalah data klimatologi harian dari BMKG Kediri periode Januari–Oktober 2024, dengan variabel suhu rata-rata (T_{avg}), kelembapan rata-rata (RH_{avg}), lama penyinaran matahari (S_s), kecepatan angin rata-rata (ff_{avg}), serta curah hujan harian (RR) sebagai target.

Tahapan penelitian meliputi pengumpulan data, praproses (standarisasi format, interpolasi data hilang, dan transformasi logaritmik), pemodelan, evaluasi, serta analisis hasil. Algoritma yang digunakan adalah Random Forest (RF) untuk mengelola hubungan non-linear, dan Vector Autoregression (VAR) untuk menganalisis keterkaitan temporal deret waktu. Evaluasi model dilakukan menggunakan Mean Squared Error (MSE) dengan skema pembagian data latih 80% dan data uji 20%.

Rancangan ini dirancang untuk membandingkan kinerja RF dan VAR secara objektif dalam memprediksi curah hujan harian di Kediri Raya.

HASIL DAN PEMBAHASAN

Penelitian ini mengikuti tahapan sistematis, mulai dari pengumpulan, pembersihan, hingga analisis data untuk membangun model prediksi curah hujan yang akurat. Data klimatologi harian dari BMKG Kota Kediri (Januari–Oktober 2024), mencakup curah hujan, suhu rata-rata, kelembapan, dan tekanan udara. Frekuensi data harian memungkinkan analisis lebih detail terhadap pola musim hujan dan kemarau.

Proses praproses dilakukan dengan interpolasi linier untuk menangani data hilang, serta visualisasi menggunakan Matplotlib dan Seaborn untuk mengidentifikasi tren dan anomali. Data dikonversi dari Excel ke format CSV agar kompatibel dengan pustaka Python seperti Pandas.

Model dibangun menggunakan dua pendekatan: Random Forest, yang efektif menangkap hubungan non-linear antar variabel, dan Vector Autoregression (VAR), yang menganalisis keterkaitan temporal dalam deret waktu. Evaluasi model menggunakan Mean Squared Error (MSE) untuk mengukur akurasi hasil prediksi.

Pustaka Python yang digunakan mencakup Pandas untuk manipulasi data, Scikit-learn untuk Random Forest, dan Statsmodels untuk VAR, mendukung seluruh proses analisis dan pemodelan secara komprehensif.

4.1 Pengambilan Data

Data penelitian diperoleh dari situs resmi BMKG untuk periode Januari hingga Oktober 2024, yang mencakup musim hujan dan kemarau di Kota Kediri. Data harian dipilih guna memperoleh resolusi yang lebih detail terhadap pola curah hujan. Format awal data berupa file Microsoft Excel kemudian dikonversi ke format CSV agar kompatibel dengan pustaka Python seperti Pandas.

Tahapan pengolahan data dilakukan secara sistematis, dimulai dari proses standarisasi format tabel dengan menghapus bagian-bagian yang tidak konsisten. Selanjutnya, nilai-nilai yang hilang dan anomali ditangani menggunakan teknik interpolasi. Setelah itu, dilakukan visualisasi awal untuk mengidentifikasi distribusi data dan potensi keberadaan outlier.

4.1.1 Import Library

Impor pustaka dilakukan sebagai tahapan awal dalam analisis dan pengembangan model prediksi. Beberapa pustaka Python digunakan untuk mendukung proses ini. Pustaka Pandas dimanfaatkan untuk mengelola dan memanipulasi

data dalam bentuk DataFrame, yang merupakan format utama dalam pengolahan data tabular. Selanjutnya, Statsmodels digunakan untuk membangun model Generalized Linear Model (GLM) dan Vector Autoregression (VAR), guna menganalisis hubungan dinamis antar variabel klimatologi. Untuk keperluan pembagian data latih dan data uji serta evaluasi kinerja model menggunakan metrik mean squared error, digunakan pustaka Scikit-learn. Terakhir, algoritma RandomForestRegressor diimplementasikan sebagai inti dari model prediksi berbasis metode Random Forest. Proses Import Library Berdasarkan Metode :

a. Random Forest (RF)

```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import mean_squared_error
```

Gambar 2. Proses Import Library untuk Implementasi RF

Menampilkan proses impor pustaka yang dibutuhkan untuk membangun model Random Forest. Selain Pandas dan Scikit-learn, pustaka RandomForestRegressor digunakan sebagai inti dari metode ensemble dalam pemodelan prediktif.

b. Vector Autoregression (VAR)

```
import pandas as pd
from sklearn.model_selection import train_test_split
from statsmodels.tsa.api import VAR
from sklearn.metrics import mean_squared_error
```

Gambar 3. Proses Import Library untuk Implementasi VAR

Menampilkan tahapan awal dalam pemanggilan pustaka untuk membangun model VAR. Library Statsmodels dimanfaatkan untuk mengevaluasi keterkaitan antar variabel deret waktu, sementara mean_squared_error digunakan sebagai alat ukur kinerja model.

4.1.2 Membaca Dataset dari File CSV

```
# load dataset dari file csv
file_path = ('C:\Users\DIAPENDIX\Documents\datasetkripsi.csv') # Ganti dengan path file CSV Anda
data = pd.read_csv(file_path)
```

Gambar 4. Proses Membaca Dataset menggunakan Pandas dari File CSV

Menjelaskan tahap membaca dataset CSV menggunakan Pandas, dimulai dari penentuan path file hingga pemuatan data ke dalam DataFrame dengan fungsi `pd.read_csv()`, sebagai dasar pengolahan data selanjutnya.

4.1.3 Menampilkan Nama Kolom dalam Dataset

```
# Tampilkan nama-nama kolom
print("Nama-nama kolom dalam dataset:")
print(data.columns.tolist())
```

Gambar 5. Proses Menampilkan Nama Kolom pada Dataset yang Dibaca

Ilustrasi ini menunjukkan cara menampilkan nama-nama kolom pada dataset menggunakan Pandas. Fungsi `print()` digunakan sebagai penanda, sedangkan `data.columns.tolist()` mengambil daftar nama kolom untuk memastikan struktur data telah sesuai sebelum dianalisis lebih lanjut.

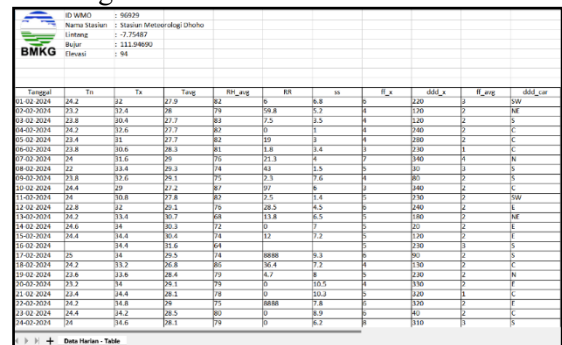
4.1.4 Memastikan Data Ter-load dengan Benar

```
# Cek beberapa baris awal dari dataset untuk memastikan data ter-load dengan benar
print("\nBeberapa baris awal dari dataset:")
print(data.head())
```

Gambar 6. Proses Verifikasi untuk Memastikan Data Ter-load dengan Benar

Menampilkan lima baris awal dataset menggunakan `data.head()` sebagai langkah verifikasi awal. Fungsi ini memastikan struktur dan isi data sesuai, serta membantu mendeteksi kolom kosong atau kesalahan pemuatan data.

4.2 Pengolahan Data



Tanggal	Tn	Tx	Tavg	RR_avg	RR	ss	ff_x	shd_x	ff_avg	shd_avg
01-02-2024	24.2	32	27.9	62	0	6.8	0	240	0	0
02-02-2024	23.3	32.8	28	59	56.8	5.7	0	230	0	0
03-02-2024	22.6	30.8	27.3	63	7.5	3.5	0	220	0	0
04-02-2024	24.2	32.6	27.7	60	0	1	0	240	0	0
05-02-2024	22.4	31	27.7	62	10	0	0	240	0	0
06-02-2024	23.8	30.6	28.3	61	1.8	1.4	0	230	0	0
07-02-2024	24	31.6	28	56	27.3	4	0	240	0	0
08-02-2024	22	31.4	26.3	74	0	1.5	0	20	0	0
09-02-2024	23.8	32.6	28.1	55	2.3	7.6	0	20	0	0
10-02-2024	24.4	29	27.2	67	0	0	0	240	0	0
11-02-2024	24	30.8	27.8	62	0.5	1.4	0	230	0	0
12-02-2024	24.8	34	28.5	59	10.5	6.5	0	240	0	0
13-02-2024	24.2	33.4	28.7	68	13.8	6.5	0	240	0	0
14-02-2024	24.6	38	30.3	72	0	0	0	250	0	0
15-02-2024	24.4	34.4	30.4	74	12	7.2	0	230	0	0
16-02-2024	24.4	31.6	28	64	0	0	0	230	0	0
17-02-2024	25	34	28.5	74	0	0	0	20	0	0
18-02-2024	24.2	33.3	28.8	65	16.4	7.2	0	230	0	0
19-02-2024	22.6	32.6	28.4	59	6.7	0	0	230	0	0
20-02-2024	21.3	34	28.1	74	0	10.1	0	230	0	0
21-02-2024	22.4	34.4	28.1	78	0	10.3	0	230	0	0
22-02-2024	24.2	34.8	28.5	74	0	0	0	230	0	0
23-02-2024	24.4	34.2	28.5	60	0	0	0	240	0	0
24-02-2024	24	34.6	28.1	79	0	6.2	0	210	0	0

Gambar 7. Dataset Penelitian

Gambar yang disajikan memperlihatkan contoh dataset harian yang diperoleh dari Stasiun Meteorologi Dhocho (BMKG). Dataset ini memuat sepuluh parameter klimatologi yang diamati secara rutin. Di antara parameter tersebut, curah hujan harian (RR) ditetapkan sebagai variabel target yang akan diprediksi. Beberapa parameter penting dalam dataset ini meliputi suhu minimum (Tn), suhu maksimum (Tx), dan suhu rata-rata (Tavg) yang semuanya diukur dalam derajat Celsius. Selain itu, terdapat kelembapan udara rata-rata (RH_avg) dalam persen, serta curah hujan harian (RR) dalam milimeter. Lama penyinaran matahari (Ss) dicatat dalam satuan jam, sementara informasi mengenai angin terdiri dari kecepatan maksimum (ff_x), arah

angin maksimum (ddd_x), kecepatan angin rata-rata (ff_avg), dan arah angin dominan (ddd_car). Dari seluruh parameter tersebut, dipilih beberapa fitur utama yang dianggap paling berkontribusi terhadap pembentukan hujan, yaitu suhu rata-rata (Tavg), kelembapan rata-rata (RH_avg), lama penyinaran matahari (Ss), dan kecepatan angin rata-rata (ff_avg). Keempat variabel ini dijadikan sebagai fitur masukan (X) dalam pemodelan, sedangkan curah hujan (RR) berfungsi sebagai variabel target (Y). Pemilihan fitur dilakukan berdasarkan pertimbangan korelasinya terhadap pola curah hujan, sehingga model prediksi dapat dibangun secara optimal dengan pendekatan relasi antara X dan Y.

a. Metode Random Forest (RF)

```
# Menyiapkan data untuk Random Forest
features = ['temperature', 'humidity', 'sunshine', 'windspeed']
X = data[features]
y = data['rainfall']
```

Gambar 8. Proses Pengolahan Data Menggunakan RF

Menjelaskan tahapan pemrosesan data dengan menggunakan model Random Forest (RF). Variabel-variabel bebas seperti suhu, kelembapan, durasi penyinaran matahari, dan kecepatan angin dijadikan fitur. Data input X diambil dari dataset, sementara variabel target y berupa curah hujan. Proses ini bertujuan untuk memastikan model mampu melakukan prediksi curah hujan secara akurat.

b. Metode Vector Autoregression (VAR)

```
# Menyiapkan data untuk VAR, fokus pada fitur dan label
# Dalam hal ini, kita menggunakan semua fitur dan 'rainfall' sebagai variabel target
data = data[['temperature', 'humidity', 'sunshine', 'windspeed', 'rainfall']]
```

Gambar 9. Proses Pengolahan Data Menggunakan VAR

Menggambarkan proses pengolahan data dengan menggunakan model Vector Autoregression (VAR). Dataset dipilih dengan subset variabel seperti suhu, kelembapan, durasi penyinaran matahari, kecepatan angin, dan curah hujan untuk dianalisis secara bersamaan. Model VAR memperhitungkan hubungan dinamis antar variabel serta efek keterlambatan waktu (lag), sehingga memungkinkan analisis interaksi variabel dalam data deret waktu.

4.3 Validasi Data

Validasi data dilakukan untuk memastikan format dataset sudah benar, mengidentifikasi nilai ekstrem, serta mempersiapkan data agar layak digunakan dalam pemodelan. Analisis awal menemukan adanya beberapa nilai nol dan data yang kurang lengkap. Untuk mengatasi hal ini, dilakukan transformasi logaritma guna memperbaiki distribusi data dan mengurangi pengaruh nilai ekstrem, sehingga distribusi data menjadi lebih mendekati

normal. Transformasi logaritma dilakukan menggunakan rumus:

$$X' = \log(X + 1) \quad (1)$$

Transformasi logaritma diterapkan untuk menstabilkan variansi data dan meningkatkan performa model prediksi. Transformasi ini dilakukan menggunakan persamaan matematis yang melibatkan fungsi logaritma terhadap nilai asli data, sehingga diperoleh nilai baru dalam skala logaritmik. Dalam konteks ini, variabel X merepresentasikan nilai asli, sedangkan X' adalah hasil transformasi logaritmik dari nilai tersebut. Proses ini dilakukan dengan menerapkan kebalikan dari transformasi logaritma menggunakan rumus berikut:

$$x_{pred} = 10^{y_{pred}} - 1 \quad (2)$$

Setelah model selesai melakukan prediksi dalam skala logaritmik, hasilnya perlu dikembalikan ke skala semula agar interpretasi lebih bermakna dalam konteks dunia nyata. Proses ini disebut sebagai inversi transformasi logaritma, yang dilakukan dengan menerapkan fungsi eksponensial terhadap nilai prediksi hasil transformasi. Nilai Y_pred mewakili hasil prediksi dalam skala logaritma, sementara X_pred adalah hasil akhir dalam bentuk aslinya setelah dilakukan inversi. Sebagai ilustrasi, jika nilai awal X = 10, setelah transformasi logaritma hasilnya menjadi sekitar 1.0414. Setelah model melakukan prediksi dalam bentuk yang telah ditransformasi, kebalikan transformasi logaritma digunakan untuk mengembalikan hasil ke skala aslinya, misalnya menjadi X_pred = 99.

Setelah validasi data selesai, dataset kemudian dibagi menjadi data latih dan data uji guna memastikan model diuji secara objektif. Berikut adalah metode pembagian data yang diterapkan dalam penelitian ini:

a. Random Forest (RF)

```
# Membagi dataset menjadi training dan testing (80% training, 20% testing)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

Gambar 10. Proses Validasi Data Menggunakan Metode GLM dan Random Forest

Pada metode ini, dataset dibagi menggunakan teknik train-test split, dengan 80% data digunakan untuk pelatihan dan 20% untuk pengujian. Data latih (X_train, y_train) digunakan untuk melatih model, sementara data uji (X_test, y_test) digunakan untuk mengevaluasi performa prediksi.

b. Metode Vector Autoregression (VAR)

```
# Membagi dataset menjadi training dan testing (80% training, 20% testing)
train_size = int(len(data) * 0.8)
train, test = data.iloc[:train_size], data.iloc[train_size:]
```

Gambar 11. Proses Validasi Data Menggunakan Metode VAR

Dalam metode VAR, pembagian dataset dilakukan berdasarkan indeks waktu. Sebanyak 80% data pertama digunakan sebagai data latih (train), sementara 20% sisanya digunakan sebagai data uji (test). Pembagian ini dilakukan untuk mempertahankan pola temporal dalam data sehingga model dapat menangkap keterkaitan antar variabel secara lebih akurat dalam analisis deret waktu.

4.4 Pemodelan Data (Modelling)

Pada tahap ini, pemodelan dilakukan dengan menggunakan dua metode utama, yaitu Random Forest (RF) dan Vector Autoregression (VAR). Masing-masing metode menerapkan pendekatan berbeda untuk mengidentifikasi pola dalam data dan diuji untuk mengetahui mana yang memberikan hasil prediksi curah hujan terbaik.

a. Metode Random Forest (RF)

Random Forest merupakan algoritma ensemble yang membangun banyak pohon keputusan (decision trees) secara bersamaan. Setiap pohon menghasilkan prediksi yang berbeda, kemudian hasil akhir diperoleh dari agregasi seluruh pohon tersebut. Metode ini efektif dalam menangani hubungan non-linear serta interaksi antar variabel. Selain itu, Random Forest dapat meningkatkan akurasi model sekaligus mengurangi risiko overfitting karena menggunakan kumpulan model pohon keputusan secara bersamaan.

```
# Membuat model Random Forest
model = RandomForestRegressor(n_estimators=100, random_state=42)
model.fit(X_train, y_train)
```

Gambar 12. Proses Pemodelan dengan Metode RF

Proses pemodelan dengan Random Forest (RF) dilakukan dengan menggunakan algoritma ensemble yang menggabungkan banyak pohon keputusan untuk meningkatkan ketepatan prediksi sekaligus meminimalkan kemungkinan overfitting. Model dibangun menggunakan RandomForestRegressor dengan parameter `n_estimators=100` untuk menetapkan jumlah pohon, serta `random_state=42` agar hasilnya dapat direproduksi. Selanjutnya, model dilatih dengan metode fit pada data pelatihan `X_train` dan `y_train`.

b. Metode Vector Autoregression (VAR)

VAR digunakan untuk mengevaluasi hubungan waktu antar variabel dalam dataset deret waktu. Metode ini tepat untuk data yang variabelnya saling memengaruhi secara temporal. Model dikembangkan dengan pustaka `statsmodels`, dan pemilihan lag dilakukan berdasarkan hasil validasi performa model. Namun, hasil VAR tidak menunjukkan peningkatan signifikan dibandingkan baseline, kemungkinan karena tidak adanya

hubungan temporal yang kuat antar variabel pada dataset.

```
# Membuat model VAR
model = VAR(train)

# Melatih model
results = model.fit(maxlags=15, ic='aic')
```

Gambar 13. Proses Pemodelan dengan metode VAR

Menampilkan tahapan pemodelan menggunakan Vector Autoregression (VAR). Model dibangun dengan fungsi `VAR(train)`, di mana data latih berisi beberapa variabel yang dianalisis untuk memahami dinamika temporal antar variabel tersebut.

Setelah model dibentuk, dilakukan pembagian dataset menggunakan fungsi `train_test_split` dari pustaka `sklearn`, dengan proporsi 80% data untuk pelatihan dan 20% untuk pengujian. Langkah ini bertujuan memastikan model tidak hanya belajar dari data yang tersedia, tetapi juga mampu membuat prediksi yang akurat terhadap data baru.

Selama proses pelatihan, model mempelajari hubungan matematis antara variabel input dan curah hujan. Kemudian, pada tahap pengujian, performa model dievaluasi menggunakan metrik Mean Squared Error (MSE). Nilai MSE yang lebih rendah menunjukkan hasil prediksi yang lebih baik.

4.5 Evaluasi

Evaluasi dilakukan untuk mengukur performa model prediksi dengan menggunakan metrik utama berupa Mean Squared Error (MSE). Metrik ini menghitung rata-rata kuadrat selisih antara nilai prediksi dan nilai aktual, sehingga cukup sensitif terhadap kesalahan yang besar. Oleh karena itu, MSE dianggap tepat untuk mengukur akurasi dalam konteks prediksi kuantitatif seperti curah hujan.

Proses evaluasi mencakup perhitungan nilai MSE dari setiap model yang dikembangkan. Hasil perhitungan ini kemudian dibandingkan dengan baseline sederhana, yaitu prediksi rata-rata dari nilai target tanpa pemodelan. Perbandingan ini bertujuan untuk mengetahui sejauh mana model mampu memberikan peningkatan prediksi yang signifikan dibandingkan metode prediksi dasar. Interpretasi hasil dilakukan dengan melihat besar kecilnya nilai MSE—semakin kecil nilainya, maka semakin baik performa model dalam melakukan prediksi. Untuk memahami implementasi evaluasi

ini, berikut adalah penjelasan kode yang digunakan pada beberapa metode pemodelan :

a. Random Forest (RF)

```
# Menghitung dan menampilkan Mean Squared Error sebagai evaluasi kinerja model
mse = mean_squared_error(y_test, predictions)
print(f"\nMean Squared Error: {mse}")
```

Gambar 14. Proses Evaluasi Hasil Prediksi Menggunakan RF

Menampilkan evaluasi prediksi menggunakan Random Forest. MSE dihitung dengan `mean_squared_error(y_test, predictions)`, lalu hasilnya ditampilkan dengan `print(f"\nMean Squared Error: {mse}")`.

b. Metode Vector Autoregression (VAR)

```
# Menghitung dan menampilkan Mean Squared Error sebagai evaluasi kinerja model
mse = mean_squared_error(comparison_df['Actual'], comparison_df['Predicted'])
print(f"\nMean Squared Error: {mse}")
```

Gambar 15. Proses Evaluasi Hasil Prediksi Menggunakan VAR

Pada model VAR, nilai aktual disimpan dalam `comparison_df['Actual']`, dan prediksi dalam `comparison_df['Predicted']`. Evaluasi dilakukan dengan menghitung MSE menggunakan `mean_squared_error()` dan menampilkan hasilnya melalui `print()` untuk analisis performa.

4.6 Model Prediksi

Model prediktif dikembangkan dengan membagi data menjadi dua bagian, yaitu sekitar delapan puluh persen untuk proses pelatihan dan sisanya digunakan sebagai data pengujian. Pada tahap pelatihan, model dibentuk untuk mempelajari pola hubungan antara variabel input (fitur/X) dan target yang ingin diprediksi (Y). Berikut implementasi prediksi dan evaluasi menggunakan beberapa metode:

a. Metode Random Forest (RF)

```
# Menampilkan nilai aktual dan prediksi
comparison_df = pd.DataFrame({'Actual': y_test, 'Predicted': predictions})
print("\nPerbandingan nilai aktual dan prediksi:")
print(comparison_df.reset_index(drop=True))

# Simpan Prediksi
predictions_df = pd.DataFrame({
    'Actual': y_test,
    'Predicted': predictions
})
predictions_df.to_csv('RF_HasilPrediksi_Results.csv', index=False)
print("\nPredicted results saved to 'RF_HasilPrediksi_Results.csv'.")
```

Gambar 16. Proses Penerapan Metode RF dalam Model Prediksi

Menampilkan hasil prediksi RF menggunakan `comparison_df` untuk membandingkan nilai aktual dan prediksi. Output disimpan ke file `RF_HasilPrediksi_Results.csv`.

b. Metode Vector Autoregression (VAR)

```
# Menampilkan nilai aktual dan prediksi
print("\nPerbandingan nilai aktual dan prediksi:")
print(comparison_df.reset_index(drop=True))

# Simpan hasil prediksi ke file csv
comparison_df.to_csv('VAR_HasilPrediksi_Results.csv', index=False)
print("\nPredicted results saved to 'VAR_HasilPrediksi_Results.csv'.")
```

Gambar 17. Proses Penerapan Metode VAR dalam Model Prediksi

Prediksi VAR ditampilkan dengan `comparison_df.reset_index(drop=True)`, kemudian disimpan ke `VAR_HasilPrediksi_Results.csv`.

4.7 Hasil Penelitian

Hasil evaluasi menunjukkan bahwa performa model dinilai berdasarkan nilai Mean Squared Error (MSE) sebagai metrik utama. Nilai MSE yang lebih kecil menandakan prediksi model lebih akurat terhadap data aktual.

Dalam penelitian ini, model Random Forest (RF) menghasilkan MSE sebesar 0.21. Meskipun RF memiliki keunggulan dalam menangani data non-linear dan kompleks, tingkat kesalahan prediksinya relatif lebih tinggi dibanding metode lain. Sebaliknya, model Vector Autoregression (VAR) menunjukkan performa lebih baik dengan MSE sebesar 0.12, yang mencerminkan kemampuannya dalam mengenali pola hubungan temporal antar variabel dalam data deret waktu.

Perbandingan ini mengindikasikan bahwa pemilihan model sangat dipengaruhi oleh karakteristik data yang digunakan. RF lebih sesuai untuk struktur data dengan interaksi non-linear, sedangkan VAR lebih optimal pada data time series dengan keterkaitan antar waktu. Dengan demikian, VAR dinilai memberikan hasil prediksi yang lebih akurat dalam konteks penelitian ini.

SIMPULAN

Penelitian ini menunjukkan bahwa metode Vector Autoregression (VAR) lebih efektif dalam menangkap dinamika temporal data curah hujan, sementara Random Forest (RF) menghasilkan prediksi yang lebih stabil. Hasil perbandingan membuktikan bahwa kombinasi keduanya berpotensi memberikan prediksi yang lebih reliabel dibandingkan penggunaan metode tunggal.

Model yang dikembangkan relevan untuk mendukung sistem informasi prediksi cuaca lokal, khususnya di sekitar Bandara Dhoho Kediri, sehingga dapat dimanfaatkan dalam mitigasi bencana hidrometeorologi, pengelolaan sumber daya air, serta strategi operasional bandara.

Penelitian selanjutnya dapat diarahkan pada pengembangan model hybrid yang mengintegrasikan RF dan VAR, serta penerapan data dengan rentang waktu lebih panjang agar hasil

prediksi semakin akurat dan adaptif terhadap perubahan iklim.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) atas penyediaan data klimatologi secara terbuka, yang memungkinkan penelitian ini dapat dilaksanakan dengan baik. Data yang diberikan memiliki peran penting dalam pengembangan model prediksi curah hujan yang lebih akurat.

DAFTAR PUSTAKA

- [1] W. Y. Ambarita, S. Dur, and S. Harleni, "Analisis Diagnostik Variabel Cuaca untuk Estimasi Pola Curah Hujan di Medan menggunakan Model Bayesian Vector Autoregressive," *J. Lebesgue J. Ilm. Pendidik. Mat. Mat. Dan Stat.*, vol. 4, no. 3, 2023, doi: 10.46306/lb.v4i3.470.
- [2] N. Ponnampuruma and L. Rajapakse, "Comparison of Time Series Forecast Models for Rainfall and Drought Prediction," in *2021 Moratuwa Engineering Research Conference (MERCon)*, 2021, pp. 626–631. doi: 10.1109/MERCon52712.2021.9525690.
- [3] P. Meghana and others, "Analysis of Neural Network Algorithm in Comparison to Multiple Linear Regression and Random Forest Algorithm," in *2024 ASU International Conference in Emerging Technologies for Sustainability and Intelligent Systems (ICETISIS)*, 2024, pp. 437–443. doi: 10.1109/ICETISIS61505.2024.10459496.
- [4] Y. Liu, M. C. Roberts, and R. Sioshansi, "A vector autoregression weather model for electricity supply and demand modeling," *J. Mod. Power Syst. Clean Energy*, vol. 6, no. 4, pp. 763–776, 2018, doi: 10.1007/s40565-017-0365-1.
- [5] R. Nasution, A. Nuraini, and F. Arifin, "Pemanfaatan Machine Learning untuk Prediksi Curah Hujan di Wilayah Tropis," *J. Inform. Dan Sains*, vol. 8, no. 1, pp. 45–53, 2022.
- [6] H. F. Widodo and others, "Evaluasi Model Ensemble Random Forest untuk Prediksi Cuaca Harian di Indonesia," *J. Sains Atmos.*, vol. 11, no. 2, pp. 89–97, 2021.
- [7] I. Prabowo and D. Hartati, "Integrasi Model Statistik dan Pembelajaran Mesin dalam Prediksi Iklim Ekstrem," *J. Teknol. Dan Sains Data*, vol. 3, no. 2, pp. 112–120, 2023.
- [8] A. Permatasari, "Penerapan Model Hybrid VAR-RF untuk Sistem Peringatan Dini Cuaca Ekstrem di Bandara," *J. Meteorol. Trop. Indones.*, vol. 10, no. 1, pp. 55–65, 2024.
- [9] G. de J. Machado-Guillén, J. M. Cruz-Duarte, S. E. Conant-Pablos, and K. Filus, "Harnessing Machine Learning for Reliable Weather Forecasting: Meteorological Impact on Sustainable Energy in Monterrey," in *2024 IEEE Congress on Evolutionary Computation (CEC)*, 2024, pp. 1–8. doi: 10.1109/CEC60901.2024.10611993.
- [10] M. M. Soe, "Rainfall Prediction using Regression Model," in *2023 IEEE Conference on Computer Applications (ICCA)*, 2023, pp. 113–117. doi: 10.1109/ICCA51723.2023.10182116.
- [11] A. T. Sari and B. Wijaya, "Analisis Prediksi Curah Hujan Menggunakan Metode Random Forest pada Wilayah Jawa Timur," *J. Tek. Inform. Dan Sist. Inf. JTEKSIS*, vol. 6, no. 2, pp. 45–52, 2023.
- [12] D. Prasetyo and R. Susanto, "Implementasi Vector Autoregression untuk Prediksi Cuaca di Wilayah Tropis," *J. Tek. Inform. Dan Sist. Inf. JTEKSIS*, vol. 7, no. 1, pp. 15–24, 2024.
- [13] E. Wibowo and M. Hartono, "Integrasi Model Statistik dan Pembelajaran Mesin untuk Sistem Peringatan Dini Cuaca Ekstrem di Indonesia," *J. Tek. Inform. Dan Sist. Inf. JTEKSIS*, vol. 5, no. 3, pp. 30–40, 2022.
- [14] M. F. Hasan, R. R. Dewi, and T. Sari, "Pemodelan Prediksi Curah Hujan Menggunakan Metode Support Vector Machine di Wilayah Tropis," *J. Teknol. Dan Sains Data*, vol. 4, no. 1, pp. 23–31, 2024.
- [15] A. Kurniawan and L. Hartono, "Analisis Perbandingan Metode Machine Learning untuk Prediksi Curah Hujan Harian di Indonesia," *J. Inform. Dan Sains*, vol. 9, no. 2, pp. 75–83, 2024.