

# KINERJA ALGORITMA PSO RANDOM FOREST VERSUS METODE KONVENSIONAL: IMPUTASI HUJAN HARIAN JAKARTA

## *PERFORMANCE OF PSO RANDOM FOREST VERSUS CONVENTIONAL METHODS: JAKARTA DAILY RAINFALL IMPUTATION*

<sup>1</sup>Ade Suhendar Sutisna. <sup>2</sup>Aep Saepuloh

<sup>1,2</sup> Program Studi Teknik Sipil, Universitas Nusa Megarkencana

<sup>1</sup>adesuhendars@unmeka.ac.id, <sup>2</sup>aepsaepuloh@unmeka.ac.id

### **Abstrak**

Penelitian ini bertujuan untuk menguji efektivitas algoritma metaheuristik dan machine learning dalam melakukan imputasi data hujan harian yang hilang, dengan fokus pada wilayah Jakarta. Kesenjangan data pada pencatatan curah hujan harian merupakan tantangan dalam analisis hidrologi dan prakiraan cuaca. Metode penelitian mencakup pengumpulan data curah hujan harian dari tiga stasiun BMKG di Jakarta (Halim Perdanakusuma, Soekarno Hatta, dan Kemayoran) untuk periode 1995-2021. Identifikasi menunjukkan adanya missing data yang signifikan: 459 data di Soekarno-Hatta, 262 di Kemayoran, dan 5694 di Halim Perdanakusuma. Algoritma Random Forest dengan optimasi Particle Swarm Optimization (PSO) dikembangkan untuk memperbaiki data hujan di Stasiun Halim Perdanakusuma, menggunakan data dari Soekarno-Hatta dan Kemayoran sebagai input. Kinerja algoritma ini dibandingkan dengan metode konvensional, yaitu metode aritmatika dan resiprok. Hasil penelitian menunjukkan bahwa metode PSO-Random Forest memiliki kemampuan yang lebih baik dalam mengimputasi data yang hilang dibandingkan dengan metode konvensional, dilihat dari nilai error yang lebih rendah dan korelasi yang lebih tinggi dengan data observasi. Penelitian ini memberikan kontribusi signifikan dalam peningkatan kualitas data curah hujan, yang sangat penting untuk keperluan analisis hidrologi, mitigasi bencana, dan perencanaan sumber daya air di wilayah Jakarta.

**Kata kunci:** Imputasi Data, Curah Hujan, Random Forest, PSO, Metode Konvensional, Jakarta.

### **Abstract**

*This study aims to examine the effectiveness of metaheuristic algorithms and machine learning in imputing missing daily rainfall data, focusing on the Jakarta region. Data gaps in daily rainfall records pose a significant challenge in hydrological analysis and weather forecasting. The research methodology includes collecting daily rainfall data from three BMKG (Meteorology, Climatology, and Geophysical Agency) stations in Jakarta—Halim Perdanakusuma, Soekarno-Hatta, and Kemayoran—for the period of 1995–2021. The identification process revealed significant missing data: 459 records at Soekarno-Hatta, 262 at Kemayoran, and 5,694 at Halim Perdanakusuma. A Random Forest algorithm optimized with Particle Swarm Optimization (PSO) was developed to reconstruct the rainfall data at Halim Perdanakusuma Station, utilizing data from Soekarno-Hatta and Kemayoran as inputs. The performance of this algorithm was compared with conventional methods, specifically the arithmetic and reciprocal methods. The results show that the PSO-Random Forest method demonstrates a better capability in imputing missing data compared to conventional methods, as evidenced by lower error values and a higher correlation with observational data. This research provides a significant contribution to improving the quality of rainfall data, which is crucial for hydrological analysis, disaster mitigation, and water resource planning in the Jakarta region.*

**Keywords:** Data Imputation, Rainfall, Random Forest, PSO, Conventional Methods, Jakarta.

## PENDAHULUAN

Curah hujan merupakan elemen krusial dalam siklus hidrologi dan memiliki dampak signifikan terhadap berbagai aspek kehidupan, mulai dari pertanian, pengelolaan sumber daya air, hingga mitigasi bencana (Rizky, 2019). Ketersediaan data curah hujan yang akurat dan lengkap sangat penting untuk keperluan analisis hidrologi, prakiraan cuaca, dan perencanaan tata ruang (Kurniawan, Hadiani, & Setiono, 2017). Namun, data curah hujan seringkali tidak lengkap akibat berbagai faktor seperti kerusakan alat, gangguan teknis, atau masalah operasional (Nugroho & Haris, 2024). Kesenjangan data ini dapat mengganggu analisis dan pengambilan keputusan yang tepat. Oleh karena itu, diperlukan metode yang efektif untuk melakukan imputasi atau pengisian data yang hilang.

Berbagai metode telah dikembangkan untuk mengatasi masalah missing data pada data curah hujan. Metode konvensional seperti metode rata-rata aritmatika, metode normal rasio, dan metode resiprok sering digunakan karena kesederhanaannya (Prawaka, Zakaria, & Tugiono, 2016; Hilwan, Manyuk, & Darmayanti, 2013; Rizky, 2019; Sekretariat Pusat Hathi, 2018). Namun, metode-metode ini mungkin tidak selalu memberikan hasil yang optimal, terutama jika terdapat variabilitas spasial dan temporal yang tinggi pada data curah hujan. Perkembangan teknologi machine learning (ML) dan algoritma metaheuristik menawarkan solusi alternatif yang lebih canggih dan berpotensi menghasilkan imputasi yang lebih akurat.

Algoritma *Random Forest* (RF) adalah salah satu metode ML yang populer digunakan dalam berbagai aplikasi, termasuk prediksi curah hujan (Maulana & Riksakomara, 2017; Tyralis, Papacharalampous, & Langousis, 2019). RF mampu menangani data berdimensi tinggi dan non-linearitas, serta memberikan estimasi yang robust terhadap outlier. Algoritma metaheuristik, seperti Particle Swarm Optimization (PSO), dapat digunakan

untuk mengoptimalkan parameter-parameter dalam model ML, sehingga meningkatkan kinerja prediksi (Penalun, Hermawan, & Avianto, 2023). PSO adalah teknik optimasi yang terinspirasi dari perilaku kawanan burung atau ikan, yang efektif dalam mencari solusi optimal dalam ruang pencarian yang kompleks.

Penelitian ini bertujuan untuk membandingkan efektivitas algoritma PSO-RF dengan metode konvensional dalam melakukan imputasi data hujan harian di wilayah Jakarta. Studi kasus ini akan menggunakan data dari tiga stasiun BMKG yang terletak di Halim Perdanakusuma, Soekarno Hatta, dan Kemayoran. Hasil penelitian diharapkan dapat memberikan kontribusi dalam meningkatkan kualitas data curah hujan dan memberikan rekomendasi metode imputasi yang paling sesuai untuk wilayah Jakarta.

## TINJAUAN PUSTAKA

### Data Hujan dan Imputasi

Data curah hujan merupakan komponen penting dalam analisis hidrologi, perencanaan sumber daya air, dan mitigasi bencana. Ketersediaan data yang lengkap dan akurat sangat krusial untuk berbagai keperluan seperti analisis frekuensi, peramalan banjir, dan pengelolaan irigasi (Kurniawan, Hadiani, & Setiono, 2017; Rizky, 2019). Namun, data curah hujan seringkali tidak lengkap karena berbagai alasan, termasuk kerusakan alat ukur, masalah teknis, atau kendala operasional (Nugroho & Haris, 2024). Kondisi ini memerlukan metode untuk mengisi data yang hilang (imputasi).

Beberapa metode konvensional yang umum digunakan untuk imputasi data hujan meliputi metode rata-rata aritmatika, metode normal rasio, dan metode resiprok (Prawaka, Zakaria, & Tugiono, 2016; Hilwan, Manyuk, & Darmayanti, 2013; Rizky, 2019; Sekretariat Pusat Hathi, 2018). Metode rata-rata aritmatika menggunakan rata-rata dari data

pada stasiun terdekat untuk mengisi data yang hilang. Metode normal rasio memperhitungkan perbedaan nilai rata-rata curah hujan antara stasiun yang akan diisi dengan stasiun referensi. Metode resiprok menggunakan kombinasi data dari beberapa stasiun dengan mempertimbangkan jarak dan korelasi antar stasiun (Kurniawan, Hadiani, & Setiono, 2017; Rizky, 2019). Metode-metode ini relatif mudah diterapkan, tetapi kinerjanya dapat bervariasi tergantung pada karakteristik data dan kondisi geografis.

Penelitian sebelumnya telah mengeksplorasi berbagai metode imputasi. Risnayah (2024) menggunakan metode *Last Observation Carried Forward* (LOCF) dan rata-rata silang untuk mengisi data curah hujan harian dari ARG (*Automatic Rain Gauges*). Hasil penelitian menunjukkan bahwa metode tersebut mampu mengurangi persentase data hilang dan menghasilkan data yang akurat. Wattimena dkk. (2025) membandingkan metode imputasi rata-rata dan interpolasi dalam model *Long Short-Term Memory* (LSTM) untuk prediksi curah hujan. Hasilnya menunjukkan bahwa interpolasi memberikan kinerja yang sedikit lebih baik dibandingkan dengan imputasi rata-rata.

### **Machine Learning dalam Pengolahan Data Hujan**

Perkembangan machine learning (ML) telah membuka peluang baru dalam pengolahan data hujan, termasuk imputasi dan prediksi. Algoritma ML mampu menangani data yang kompleks dan non-linear, serta memberikan hasil yang lebih akurat dibandingkan dengan metode konvensional (Tyrallis, Papacharalampous, & Langousis, 2019). Beberapa algoritma ML yang populer digunakan dalam bidang hidrologi antara lain *Random Forest* (RF), *Support Vector Machine* (SVM), dan *Artificial Neural Network* (ANN).

*Random Forest* (RF) adalah algoritma *ensemble learning* yang terdiri dari banyak *decision tree*. RF mampu menangani data

berdimensi tinggi, mengatasi masalah *overfitting*, dan memberikan estimasi yang *robust* (Maulana & Riksakomara, 2017; Tyrallis, Papacharalampous, & Langousis, 2019). Penelitian oleh Maulana dan Riksakomara (2017) menggunakan RF untuk memprediksi intensitas hujan di Surabaya. Hasilnya menunjukkan bahwa RF mampu mengklasifikasi intensitas hujan dengan akurasi yang baik. Penalun, Hermawan, & Avianto (2023) membandingkan RF dengan *Support Vector Regression* (SVR) dalam memprediksi laju penguapan. Hasilnya menunjukkan bahwa SVR memiliki kinerja yang sedikit lebih baik, tetapi RF tetap merupakan pilihan yang baik.

*Artificial Neural Network* (ANN) adalah model yang terinspirasi dari struktur dan fungsi otak manusia. ANN mampu mempelajari pola-pola kompleks dalam data dan memberikan prediksi yang akurat (Panggabean, Sihombing, & Aruan, 2021). Panggabean, Sihombing, & Aruan (2021) menggunakan ANN untuk memprediksi curah hujan dan kecepatan angin. Hasil penelitian menunjukkan bahwa ANN memberikan hasil yang baik dalam memprediksi kedua variabel tersebut. Nugroho dan Haris (2024) menggunakan *Long Short-Term Memory* (LSTM), turunan dari *Recurrent Neural Network* (RNN), untuk menangani *missing values* pada data cuaca. Metode imputasi yang terbaik adalah KNN dipadukan dengan metode *Bidirectional LSTM*.

### **Optimasi Metaheuristik**

Algoritma metaheuristik adalah teknik optimasi yang digunakan untuk mencari solusi optimal dalam ruang pencarian yang kompleks. Algoritma ini sering digunakan untuk mengoptimalkan parameter-parameter dalam model ML, sehingga meningkatkan kinerja prediksi. Particle Swarm Optimization (PSO) adalah salah satu algoritma metaheuristik yang populer digunakan (Penalun, Hermawan, & Avianto, 2023). PSO

terinspirasi dari perilaku kawanan burung atau ikan, di mana setiap partikel (solusi) bergerak dalam ruang pencarian dengan mempertimbangkan posisi terbaik yang pernah dicapai oleh dirinya sendiri dan oleh seluruh kawanan. Dalam konteks imputasi data hujan, PSO dapat digunakan untuk mengoptimalkan parameter-parameter dalam model ML seperti RF. Dengan mengoptimalkan parameter-parameter tersebut, model RF dapat menghasilkan prediksi yang lebih akurat untuk mengisi data yang hilang. Penggunaan PSO dalam optimasi model ML telah terbukti efektif dalam berbagai aplikasi, termasuk prediksi curah hujan dan pengelolaan sumber daya air (Penalun, Hermawan, & Avianto, 2023).

## METODOLOGI PENELITIAN

### Data

Penelitian ini menggunakan data curah hujan harian dari tiga stasiun pengamatan BMKG di wilayah Jakarta: Stasiun Halim Perdanakusuma, Stasiun Soekarno Hatta, dan Stasiun Kemayoran. Data yang digunakan mencakup periode dari tahun 1995 hingga 2021. Data curah hujan harian diperoleh dari database BMKG dan telah melalui proses validasi awal untuk memastikan kualitas data.

Tahap awal penelitian melibatkan identifikasi dan analisis data yang hilang (missing data) pada setiap stasiun. Jumlah data yang hilang dicatat untuk masing-masing stasiun untuk memberikan gambaran mengenai tingkat kesenjangan data. Analisis

ini penting untuk memahami tantangan dalam imputasi data dan sebagai dasar untuk evaluasi kinerja metode imputasi. Tabel 1 meringkas jumlah data yang hilang untuk setiap stasiun. Terlihat bahwa Stasiun Halim Perdanakusuma memiliki jumlah data yang hilang paling banyak, yang menjadi fokus utama dalam implementasi dan evaluasi metode imputasi.

### Metode Imputasi

Penelitian ini membandingkan kinerja tiga metode imputasi: (1) metode PSO-Random Forest, (2) metode rata-rata aritmatika, dan (3) metode resiprok. Metode PSO-Random Forest menggunakan algoritma Random Forest sebagai model dasar untuk imputasi, dengan parameter-parameter model dioptimalkan menggunakan algoritma Particle Swarm Optimization (PSO).

PSO-Random Forest: Metode ini terdiri dari dua tahap utama: (1) pelatihan model Random Forest dan (2) optimasi parameter menggunakan PSO. Model Random Forest dilatih menggunakan data curah hujan dari Stasiun Soekarno Hatta dan Kemayoran sebagai variabel input untuk memprediksi curah hujan di Stasiun Halim Perdanakusuma. PSO digunakan untuk mengoptimalkan parameter-parameter penting dalam model Random Forest, seperti jumlah pohon (number of trees) dan kedalaman maksimum pohon (maximum depth). Fungsi objektif dalam PSO adalah meminimalkan nilai Root Mean Squared Error (RMSE) antara nilai prediksi dan nilai observasi pada data validasi.

**Tabel 1. Ringkasan Data Missing Values**

| Stasiun        | Periode Data | Jumlah Data Harian | Jumlah <i>Missing</i> Data |
|----------------|--------------|--------------------|----------------------------|
| Halim PK       | 1995-2021    | 9861               | 5694                       |
| Soekarno Hatta | 1995-2021    | 9861               | 459                        |
| Kemayoran      | 1995-2021    | 9861               | 262                        |

Metode Rata-rata Aritmatika: Metode ini menggunakan rata-rata curah hujan harian dari stasiun Soekarno Hatta dan Kemayoran (jika tersedia) untuk mengisi data yang hilang di Stasiun Halim Perdanakusuma. Jika data dari salah satu stasiun tidak tersedia, maka hanya data dari stasiun yang tersedia yang digunakan. Jika kedua stasiun tidak memiliki data, maka data yang hilang tidak diisi.

Metode Resiprok: secara konseptual, metode resiprok hampir sama dengan Metode Normal Ratio karena keduanya memperhitungkan stasiun-stasiun curah hujan yang saling berdekatan. Namun, letak perbedaan utamanya ada pada variabel yang digunakan; metode resiprok menggunakan jarak stasiun terdekat ke stasiun yang datanya hilang sebagai variabel utama. Dalam perhitungannya, jarak antar stasiun bertindak sebagai faktor pembagi (pembobot), yang menandakan bahwa persebaran hujan sangat dipengaruhi oleh faktor jarak. Semakin dekat jarak stasiun referensi terhadap stasiun yang datanya hilang, maka semakin besar bobot yang diberikan pada data dari stasiun referensi tersebut. Rumus yang digunakan adalah:

$$P_x = \frac{\sum_{i=1}^n \frac{P_i}{L_i^2}}{\sum_{i=1}^n \frac{1}{L_i^2}}$$

di mana:

- P<sub>x</sub> : Curah hujan yang hilang atau sedang dicari di stasiun X (biasanya dalam satuan mm).
- P<sub>i</sub> : Data curah hujan terukur di stasiun referensi/sekitarnya pada periode yang sama (mm).
- L<sub>i</sub> : Jarak antara stasiun X (yang datanya hilang) dengan stasiun referensi ke-i (bisa dalam meter atau km).
- n : Jumlah stasiun referensi yang digunakan dalam perhitungan

## HASIL DAN PEMBAHASAN

### Identifikasi Missing Data

Analisis awal terhadap data curah hujan harian menunjukkan adanya jumlah missing data yang signifikan pada ketiga stasiun. Stasiun Halim Perdanakusuma memiliki jumlah data hilang paling banyak (5694 data), diikuti oleh Stasiun Soekarno Hatta (459 data) dan Stasiun Kemayoran (262 data), seperti yang ditunjukkan pada Tabel 1. Kesenjangan data ini menjadi tantangan utama dalam penelitian ini. Data yang hilang ini disebabkan oleh berbagai faktor, termasuk kerusakan alat, pemeliharaan rutin, dan masalah transmisi data. Kondisi ini sejalan dengan penelitian sebelumnya yang menunjukkan bahwa missing data merupakan masalah umum dalam data curah hujan (Nugroho & Haris, 2024).

### Kinerja Metode Imputasi

Kinerja ketiga metode imputasi dievaluasi menggunakan metrik RMSE, MAE, dan koefisien korelasi (r). Hasil evaluasi disajikan pada Tabel 2. Tabel 2 menunjukkan bahwa metode PSO-Random Forest menghasilkan kinerja terbaik dibandingkan dengan metode lainnya. Metode ini memiliki nilai RMSE dan MAE terendah (240.93 mm dan 8.50 mm, berturut-turut) serta koefisien korelasi tertinggi (0.69). Hal ini mengindikasikan bahwa metode PSO-Random Forest memberikan imputasi yang paling akurat dibandingkan dengan metode rata-rata aritmatika dan metode resiprok. Metode rata-rata aritmatika menunjukkan kinerja terburuk dengan nilai RMSE tertinggi (299.17 mm) dan koefisien korelasi terendah (0.53). Metode resiprok memiliki kinerja yang lebih baik daripada metode rata-rata aritmatika, tetapi masih di bawah metode PSO-Random Forest. Hasil penelitian ini konsisten dengan penelitian sebelumnya yang menunjukkan bahwa metode berbasis machine learning mampu memberikan kinerja yang lebih baik dalam imputasi data dibandingkan dengan metode konvensional (Wattimena dkk., 2025; Nugroho & Haris, 2024).

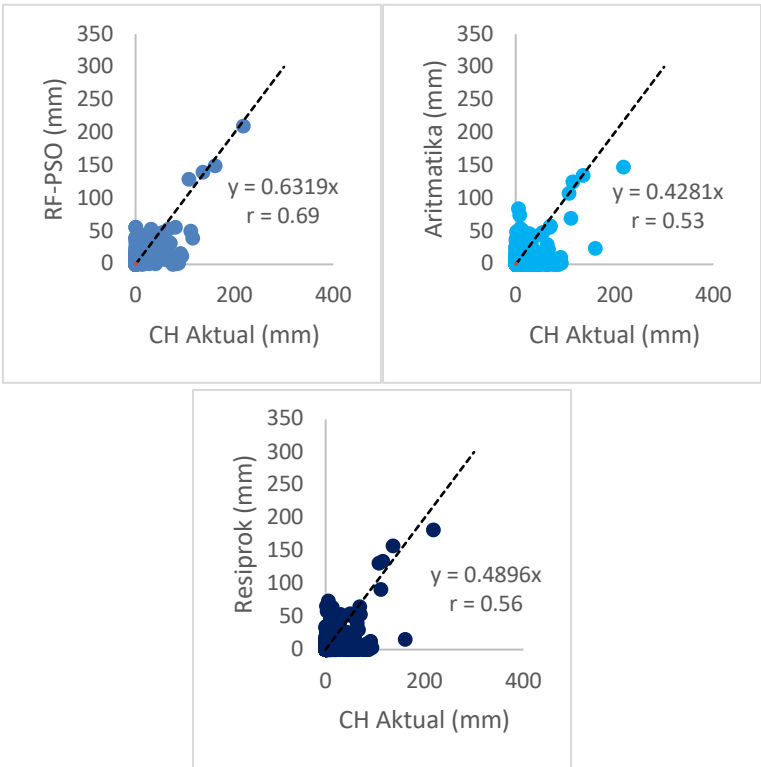
Keunggulan PSO-Random Forest disebabkan oleh kemampuannya dalam menangani non-linearitas dan kompleksitas data curah hujan, serta kemampuan PSO dalam mengoptimalkan parameter-parameter model. Penggunaan PSO dalam mengoptimalkan parameter model ML juga terbukti meningkatkan akurasi prediksi (Penalun, Hermawan, & Avianto, 2023).

### Visualisasi Hasil

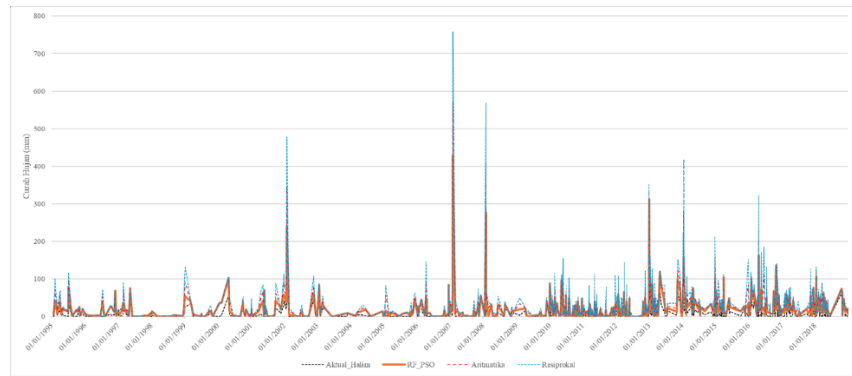
Untuk memberikan gambaran visual mengenai kinerja metode imputasi, grafik perbandingan antara nilai curah hujan yang diimputasi dengan nilai observasi dibuat. Gambar 1 menunjukkan contoh hasil imputasi menggunakan metode *PSO-Random Forest*, metode rata-rata aritmatika, dan metode resiprok.

**Tabel 2. Perbandingan Kinerja Metode Imputasi**

| Metode Imputasi      | RMSE (mm) | MAE (mm) | Koefisien Korelasi (r) |
|----------------------|-----------|----------|------------------------|
| PSO-Random Forest    | 240.93    | 8.50     | 0.69                   |
| Rata-rata Aritmatika | 299.17    | 8.81     | 0.53                   |
| Resiprok             | 291.44    | 8.67     | 0.56                   |



**Gambar 1. Korelasi data Hasil Imputasi dengan Data Observasi**



**Gambar 2. Perbandingan Hasil Imputasi dengan Data Observasi**

Gambar 2 menampilkan grafik yang membandingkan nilai curah hujan yang diimputasi (garis berwarna) dengan nilai observasi (garis hitam) untuk periode waktu tertentu. Grafik tersebut menunjukkan bahwa imputasi menggunakan metode *PSO-Random Forest* (garis oranye) lebih mendekati nilai observasi dibandingkan dengan metode rata-rata aritmatika (garis oranye putus-putus) dan metode resiprok (garis biru putus-putus). Perbedaan ini terutama terlihat pada saat terjadi curah hujan yang ekstrem, di mana metode *PSO-Random Forest* mampu menangkap variasi data dengan lebih baik.

Hasil penelitian ini menunjukkan bahwa metode *PSO-Random Forest* memberikan kinerja yang lebih baik dalam imputasi data hujan harian dibandingkan dengan metode konvensional (rata-rata aritmatika dan resiprok) di wilayah Jakarta.

Keunggulan ini sejalan dengan penelitian sebelumnya yang menunjukkan bahwa metode berbasis machine learning memiliki potensi yang lebih besar dalam pengolahan data hidrologi (Wattimena dkk., 2025; Nugroho & Haris, 2024; Longman, Newman, Giambelluca, & Lucas, 2020). Kemampuan *PSO-Random Forest* dalam menangani non-linearitas dan kompleksitas data curah hujan, serta kemampuannya dalam mengoptimalkan parameter-parameter model, menjadi faktor kunci dalam peningkatan akurasi imputasi.

Metode rata-rata aritmatika, yang merupakan metode paling sederhana, menunjukkan kinerja terburuk dalam penelitian ini. Hal ini kemungkinan disebabkan oleh asumsi linieritas yang mendasarinya, yang tidak selalu sesuai dengan karakteristik data curah hujan yang kompleks. Metode resiprok, yang mempertimbangkan korelasi antar stasiun, menunjukkan kinerja yang lebih baik daripada metode rata-rata aritmatika, tetapi masih di bawah metode *PSO-Random Forest*. Peningkatan kinerja ini menunjukkan pentingnya mempertimbangkan hubungan spasial antar stasiun dalam imputasi data.

Perbandingan hasil penelitian ini dengan penelitian lain memberikan wawasan lebih lanjut. Risnayah (2024) menggunakan metode LOCF dan rata-rata silang untuk mengisi data curah hujan ARG. Hasil penelitian tersebut menunjukkan bahwa metode LOCF dan rata-rata silang efektif dalam mengurangi missing data dan menghasilkan data yang akurat. Namun, penelitian ini menggunakan data dengan karakteristik yang berbeda (data ARG), dan metode yang digunakan juga berbeda. Wattimena dkk. (2025) membandingkan metode imputasi rata-rata dan interpolasi dalam model LSTM. Hasil penelitian mereka menunjukkan bahwa interpolasi memiliki kinerja yang lebih baik. Perbedaan hasil ini dapat disebabkan oleh perbedaan metode, data, dan wilayah studi. Penelitian ini menggunakan

PSO-Random Forest yang terbukti lebih efektif dibandingkan dengan metode konvensional dan sejalan dengan penelitian lain yang menunjukkan keunggulan ML.

Implikasi dari penelitian ini sangat penting dalam konteks pengelolaan sumber daya air, mitigasi bencana, dan perencanaan tata ruang di wilayah Jakarta. Data curah hujan yang akurat dan lengkap merupakan dasar yang penting untuk berbagai analisis dan pengambilan keputusan. Metode PSO-Random Forest dapat digunakan untuk mengisi data yang hilang, sehingga meningkatkan kualitas data curah hujan dan memungkinkan analisis yang lebih komprehensif. Penggunaan data yang lebih baik akan meningkatkan keakuratan model hidrologi, prakiraan banjir, dan perencanaan mitigasi bencana. Selain itu, hasil penelitian ini dapat digunakan untuk mengembangkan sistem peringatan dini banjir yang lebih efektif.

Meskipun demikian, penelitian ini memiliki beberapa keterbatasan. Pertama, penelitian ini hanya menggunakan data dari tiga stasiun BMKG. Peningkatan jumlah stasiun dapat meningkatkan akurasi imputasi dan memberikan gambaran yang lebih komprehensif mengenai pola curah hujan di wilayah Jakarta. Kedua, penelitian ini hanya menggunakan satu algoritma metaheuristik (PSO). Penggunaan algoritma metaheuristik lain, seperti algoritma genetika atau ant colony optimization, dapat memberikan hasil yang berbeda. Ketiga, penelitian ini tidak mempertimbangkan faktor-faktor lain yang dapat mempengaruhi curah hujan, seperti data suhu, kelembaban, dan kecepatan angin. Penambahan variabel-variabel ini dapat meningkatkan akurasi imputasi.

## SIMPULAN

Penelitian ini menyimpulkan bahwa metode PSO-Random Forest merupakan metode yang efektif untuk imputasi data hujan harian di wilayah Jakarta. Metode ini

memberikan kinerja yang lebih baik dibandingkan dengan metode konvensional (rata-rata aritmatika dan resiprok), dilihat dari nilai RMSE, MAE, dan koefisien korelasi. Keunggulan PSO-Random Forest menunjukkan potensi penggunaan algoritma metaheuristik dan machine learning dalam meningkatkan kualitas data curah hujan. Hasil penelitian ini memberikan kontribusi yang signifikan dalam peningkatan kualitas data curah hujan, yang sangat penting untuk keperluan analisis hidrologi, mitigasi bencana, dan perencanaan sumber daya air di wilayah Jakarta.

## DAFTAR PUSTAKA

- Kurniawan, R. D., Hadiani, R. R. R., & Setiono, S. (2017). Mengisi Data Hujan Yang Hilang Dengan Metode Autoregressive Dan Metode Reciprocal Dengan Pengujian Debit Kala Ulang (Studi Kasus Di Das Bakalan). *Matriks Teknik Sipil*, 5(4). <http://eprints.uns.ac.id/37916/>
- Laily, Y. H., Rakhmawati, F., & Husein, I. (2023). Penerapan Metode Fuzzy Time Series-Markov Chain Dalam Peramalan Curah Hujan Sebagai Jadwal Tanaman Padi. *Jurnal Lebesgue Jurnal Ilmiah Pendidikan Matematika Matematika dan Statistika*, 4(1), 32-40. <https://doi.org/10.46306/lb.v4i1.235>
- Longman, R. J., Newman, A. J., Giambelluca, T. W., & Lucas, M. P. (2020). Characterizing the Uncertainty and Assessing the Value of Gap-Filled Daily Rainfall Data in Hawaii. *Journal of Applied Meteorology and Climatology*, 59(7), 1189-1206. <https://doi.org/10.1175/jamc-d-20-0007.1>
- Maulana, D., & Riksakomara, E. (2017). Prediksi Intensitas Hujan Kota Surabaya dengan Matlab menggunakan Teknik Random Forest dan CART (Studi Kasus Kota Surabaya). *Jurnal Teknik ITS*, 6(1), F11-F16.



- <https://doi.org/10.12962/j23373539.v6i1.21120>
- Nugroho, A. A., & Haris, M. (2024). Analisis Efektivitas Teknik Imputasi Pada Lstm Untuk Meningkatkan Kualitas Data Pada Peramalan Curah Hujan. *Jurnal Informatika dan Rekayasa Elektronik*, 7(2), 167-176.  
<https://doi.org/10.36595/jire.v7i2.1172>
- Panggabean, D. A. H., Sihombing, F. M., & Aruan, N. M. (2021). Prediksi Tinggi Curah Hujan Dan Kecepatan Angin Berdasarkan Data Cuaca Dengan Penerapan Algoritma Artificial Neural Network (ANN). *seminastika*, 3(1), 77-85.<https://doi.org/10.47002/seminastika.v3i1.237>
- Penalun, F. E., Hermawan, A., & Avianto, D. (2023). Perbandingan Random Forest Regression dan Support Vector Regression Pada Prediksi Laju Penguapan. *JURNAL FASILKOM*, 13(02), 79-88.  
<https://doi.org/10.37859/jf.v13i02.4976>
- Prawaka, F., Zakaria, A., & Tugiono, S. (2016). Analisis Data Curah Hujan yang Hilang Dengan Menggunakan Metode Normal Ratio, Inversed Square Distance, dan Cara Rata-Rata Aljabar (Studi Kasus Curah Hujan Beberapa Stasiun Hujan Daerah Bandar Lampung). *Jurnal Rekayasa Sipil dan Desain*, 4(3), 161-169.  
<https://doi.org/10.23960/jrsdd.v4i3.418>
- Risnayah, S. (2024). Penerapan Imputasi Locf Dan Cross Mean Dalam Pengisian Data Kosong Pada Curah Hujan Harian ARG. *Megasains : Buletin Meteorologi, Klimatologi, Kualitas Udara dan Geofisika*, 14(2), 170-179.  
<https://doi.org/10.46824/megasains.v14i2.138>
- Rizky, M. (2019). Pengisian Data Curah Hujan Yang Hilang Pada Das Lariang.  
<http://repository.untad.ac.id/1439/>
- Sekretariat Pusat Hathi. (2018). Kajian Pengembangan Pengisian Kekosongan Data Hujan (Studi Kasus DAS Serang, Ws Jratun Seluna – Jawa Tengah).
- Silalahi, F. S. A., Farda, N. M., & Nurjani, E. (2024). Uji Perbandingan Metode Estimasi Curah Hujan Menggunakan Satelit Himawari: Metode Konvensional dan Machine Learning. *Jurnal Geosains dan Remote Sensing*, 5(1), 17-26.  
<http://dx.doi.org/10.23960/jgrs.ft.unila.333>
- Teegavarapu, R. S. V., Aly, A. H., Pathak, C. S., Ahlquist, J. E., Fuelberg, H. E., & Hood, J. (2017). Infilling missing precipitation records using variants of spatial interpolation and data-driven methods: use of optimal weighting parameters and nearest neighbour-based corrections. *International Journal of Climatology*, 37(10), 3843-3861.  
<https://doi.org/10.1002/joc.5209>
- Tyralis, H., Papacharalampous, G., & Langousis, A. (2019). A Brief Review of Random Forests for Water Scientists and Practitioners and Their Recent History in Water Resources. *Water*, 11(5), 910.<https://doi.org/10.3390/w11050910>
- Waqas, M., & Humphries, U. W. (2024). A critical review of RNN and LSTM variants in hydrological time series predictions. *MethodsX*, 12, 102946.  
<https://doi.org/10.1016/j.mex.2024.102946>
- Wattimena, E. M. C., Taihuttu, P. D. M., Waas, D. V., Palembang, C. F., & Pattiradjawane, V. E. (2025). Optimization of LSTM Model for Rainfall Prediction in Ambon City: Comparison of Mean Imputation and Interpolation in Time Series Data Prediction. *Tensor Pure and Applied Mathematics Journal*, 6(1), 49-56