

# Analisis Data Curah Hujan Yang Kosong Menggunakan Metode Interpolasi Linier

<sup>1</sup>Beni Andriansah<sup>2</sup>Nurul Humam<sup>3</sup>Pebry Ajeng Cahyani

Mahasiswa Program Studi Ilmu Komputer Jakarta Selatan, Indonesia

Email:<sup>1</sup>[nur.humam04@gmail.com](mailto:nur.humam04@gmail.com) <sup>2</sup>[pebryac12@gmail.com](mailto:pebryac12@gmail.com)

## ABSTRAK

---

Indonesia merupakan negara beriklim tropis lembab yang mempunyai dua musim, yaitu musim hujan dan musim kemarau. Curah hujan merupakan salah satu faktor yang penting untuk melangsungkan kehidupan di Indonesia. Penelitian ini menggunakan data curah hujan yang kosong atau hilang di wilayah Jakarta menggunakan metode interpolasi linier. Dataset yang digunakan mencakup data curah hujan di Jakarta dari bulan Mei 2023 - April 2024, dengan parameter seperti suhu rata-rata, suhu maksimum dan minimum, total curah hujan, serta kecepatan angin rata-rata. Sebanyak 83% data dihapus secara sengaja untuk mensimulasikan data yang hilang. Metode interpolasi linier diterapkan menggunakan Python dan Jupyter Notebook untuk memperkirakan data yang hilang. Hasil penelitian menunjukkan bahwa metode interpolasi linier memberikan akurasi yang lebih baik dibandingkan metode interpolasi lainnya, dibuktikan dengan nilai Mean Square Error (MSE) yang lebih rendah. Penelitian ini diharapkan dapat meningkatkan akurasi prediksi curah hujan di Indonesia, yang penting untuk berbagai sektor seperti pertanian, pengelolaan sumber daya air, dan penanggulangan bencana.

---

## ABSTRACT

---

Indonesia is a country with a humid tropical climate which has two seasons, namely the rainy season and the dry season. Rainfall is an important factor for sustaining life in Indonesia. This research uses empty or missing rainfall data in the Jakarta area using the linear interpolation method. The dataset used includes rainfall data in Jakarta from May 2023 - April 2024, with parameters such as average temperature, maximum and minimum temperature, total rainfall, and average wind speed. As much as 83% of the data was deleted intentionally to include missing data. The linear interpolation method was applied using Python and Jupyter Notebook to infer missing data. The research results show that the linear interpolation method provides better accuracy than other interpolation methods, as evidenced by the lower Mean Square Error (MSE) value. This research is expected to improve the accuracy of rainfall predictions in Indonesia, which is important for various sectors such as agriculture, air resource management and disaster prevention.

## PENDAHULUAN

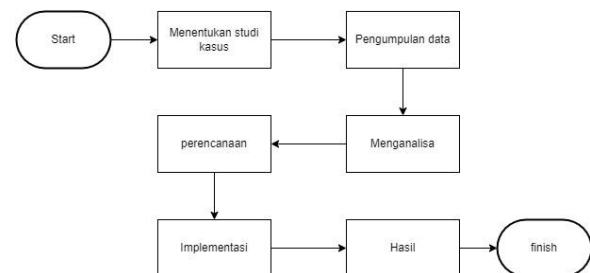
Berdasarkan letak astronomis dan geografisnya, Indonesia merupakan negara dengan iklim tropis lembab yang mempunyai dua musim, yaitu musim hujan dan musim kemarau. Secara astronomis, Indonesia terletak antara  $6^{\circ}$  Lintang Utara dan  $11^{\circ}$  Lintang Selatan serta antara  $95^{\circ}$  Bujur Timur dan  $141^{\circ}$  Bujur Timur, dan Indonesia menerima sinar matahari setiap tahunnya. Dan secara geografis, Indonesia terdapat Samudera Hindia dan Samudera Pasifik. Indonesia, juga terdapat dua barisan pegunungan muda yaitu Pegunungan Mediterania dan Lingkar Pasifik [1]. Pada iklim tropis di Indonesia, penguapan air ke atmosfer sangat tinggi sehingga curah hujan menjadi tidak menentu [2]. Curah hujan merupakan salah satu faktor yang penting di Indonesia untuk melangsungkan kehidupan, khususnya dalam mendukung berbagai sektor, seperti pertanian, pengelolaan sumber daya air, penanggulangan bencana, dan lain sebagainya.

Oleh karena itu, data curah hujan terus dicatat untuk memudahkan kelangsungan hidup masyarakat di Indonesia. Sedangkan sulit untuk mengetahui curah hujan kecuali dengan melakukan prediksi. Prediksi adalah suatu proses memperkirakan secara sistematis tentang sesuatu yang mungkin terjadi di masa yang akan datang berdasarkan data di masa lampau. Namun, seringkali data curah hujan yang diperoleh tidak lengkap atau bahkan kosong. Hal ini dapat disebabkan oleh berbagai faktor, seperti kerusakan alat pengukur, kesalahan manusia, atau keterbatasan teknologi [1]. Sehingga, prediksi curah hujan dalam waktu tertentu menjadi tidak akurat karena beberapa stasiun hujan kehilangan data.

Dalam penelitian ini, kami menggunakan metode interpolasi linier sebagai salah satu metode pendekatan secara numeris untuk mencari dataset curah hujan yang kosong dalam jangka waktu tertentu. Jangka waktu curah hujan yang digunakan dalam penelitian ini, yaitu kisaran bulan Mei 2023 - April 2024 di wilayah Jakarta. Penelitian ini mengambil data curah hujan dari meteomanz yang merupakan situs web data meteorologi. Metode interpolasi linier dalam penelitian ini diharapkan dapat memperkirakan besarnya korelasi data hasil interpolasi linier dan data hujan asli atau data lengkap sebelum adanya data yang kosong.

## METODE PENELITIAN

Dalam penelitian kali ini, penulis menggunakan code python dan jupyter notebook yang sudah dirancang sebelumnya. Metodologinya dijelaskan dibawah ini (Gambar 1).



Gambar 1. Diagram alur kerja

## Daerah Diteliti

Wilayah Jakarta tersebar di wilayah seluas 661,5 km<sup>2</sup> (255,8 mil persegi) dan berpenduduk 10.562.088 jiwa seperti yang dilaporkan pada sensus 2020. Jakarta terdiri dari lima kota administrasi: Jakarta Pusat, Jakarta Utara, Jakarta Barat, Jakarta Selatan, dan Jakarta Timur, serta satu kabupaten administrasi, yaitu Kepulauan Seribu. Jakarta Pusat adalah

pusat pemerintahan dan kota utama. Wilayah Jakarta sering mengalami banjir selama musim hujan, yang memengaruhi hampir seluruh wilayahnya. Curah hujan tahunan rata-rata di Jakarta adalah sekitar 1.800 mm. Lahan urban mencakup sebagian besar wilayah Jakarta, dengan pertanian yang sangat minim. Banjir memberikan pengaruh yang luar biasa terhadap kehidupan masyarakat Jakarta. Tahun 2020 merupakan salah satu tahun dengan curah hujan tinggi yang menyebabkan banjir besar, merusak rumah dan infrastruktur. Banjir dengan tingkat keparahan yang berbeda-beda sering terjadi di sebagian besar wilayah Jakarta dalam beberapa dekade terakhir.

## Dataset

Basis data iklim diambil dari website meteomanz.com. Dalam studi kasus ini penulis mengambil data dari daerah jakarta(96745) pada bulan Mei 2023 - April 2024 dengan jumlah 415 dataset, penulis meminimalkan dataset menjadi 12 dataset dengan hitungan bulan (tabel 1). dalam studi kasus ini penulis memprediksi dari beberapa parameter seperti Suhu rata-rata (AG), Suhu maksimum rata-rata (MMAX), Suhu maksimum tertinggi (HMAX), Suhu minimum rata-rata (MMIN), Suhu minimum terendah (LMIN), Total curah hujan (TMRP), Kecepatan Angin rata-rata (MWS). Data Curah Hujan Bulan Mei – April (Tabel 1).

Tahun	Bulan	AVG	MMAX	MMIN	HMAX	LMIN	TMRP	MWS
2023	Mei	29.5	33.2	26.97	34.6	25.8	107.5	4.0
	Juni	29.1	33.0	26.27	34.4	24.6	128.3	3.5
	Juli	29.0	32.7	25.99	36.6	24.6	6.0	3.7
	Agustus	29.0	33.0	26.15	34.8	24.0	2.0	4.4
	September	29.3	33.07	26.14	36.8	24.2	0.0	3.3
	Oktober	30.2	34.0	27.1	36.6	25.0	0.0	4.4
	November	29.5	33.3	26.5	35.6	24.4	152.4	3.9
	Desember	29.5	32.9	26.4	35.6	24.8	11.0	4.4
	Januari	28.6	32.3	25.9	35.4	24.6	311.0	5.8
	Februari	28.5	31.8	25.9	34.2	24.4	398.2	5.3
	Maret	28.6	32.1	25.9	35.4	24.0	302.0	4.8
	April	29.6	33.3	26.6	35	25.2	103.1	4.7

Tabel 1. Data Curah Hujan Bulan Mei - April

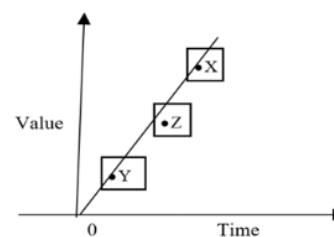
## Nilai yang Hilang

Penulis telah menghapus 83% data dari data yang telah dikumpulkan. Tujuan penghapusan data adalah untuk memeriksa perbedaan dari data asli dan data temuan. Pada dataset, penulis menghapus 2 bulan sehingga tersisa 10 bulan, yang dimana data 2 bulan tersebut akan ditemukan dalam analisa.

## Interpolasi Linier

Metode interpolasi linier merupakan salah satu teknik interpolasi yang biasa digunakan dalam pencarian data hilang atau kosong. Metode interpolasi linier dapat diasumsikan sebagai hubungan linier antara nilai yang hilang dan tidak hilang seperti yang ditunjukkan pada (Gambar 2) asumsikan data dalam angka z tidak ada, kita dapat mencari dengan data x dan data y dengan rumus berikut.

$$y = \frac{y_2 - y_1}{x_2 - x_1} (x - x_1) + y_1$$



Gambar 2. Metode Interpolasi Linier

## HASIL DAN PEMBAHASAN

### Analisis Data Awal

Penelitian ini menggunakan dataset curah hujan bulanan di Jakarta dari Mei 2023 hingga April 2024. Dataset ini mencakup tujuh parameter utama: suhu rata-rata (AVG), suhu maksimum rata-rata (MMAX), suhu maksimum tertinggi (HMAX), suhu minimum rata-rata (MMIN), suhu minimum terendah (LMIN), total curah hujan (TMRF), dan kecepatan angin rata-rata (MWS)..

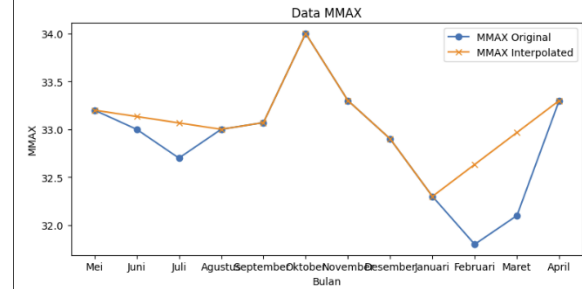
### Proses Interpolasi

Dalam penerapan ini, kami secara sengaja menghapus sebagian data untuk mensimulasikan data yang hilang. Dari dataset awal yang terdiri dari 12 bulan, kami memilih 8 bulan sebagai data yang tersedia (Mei, Agustus, September, Oktober, November, Desember, Januari, dan April). Data untuk bulan Juni, Juli, Februari, dan Maret dihapus dan kemudian diinterpolasi menggunakan metode interpolasi linier.

### Hasil Interpolasi

Setelah melakukan interpolasi, kami mendapatkan dataset yang lengkap kembali. Berikut adalah analisis perbandingan antara data asli dan hasil interpolasi untuk setiap parameter:

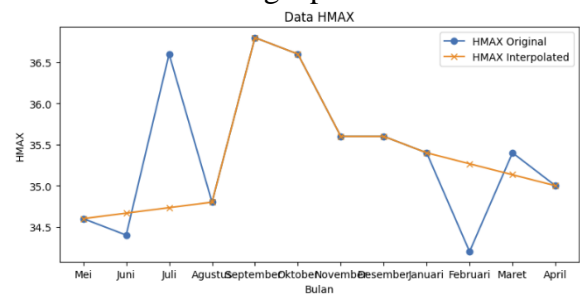
- a. Suhu Maksimum Rata-rata (MMAX)



Hasil interpolasi untuk MMAX cukup mendekati data asli. Namun, terdapat sedikit perbedaan pada bulan Juli dan Agustus, di mana interpolasi tidak sepenuhnya menangkap penurunan suhu yang terjadi pada data asli.

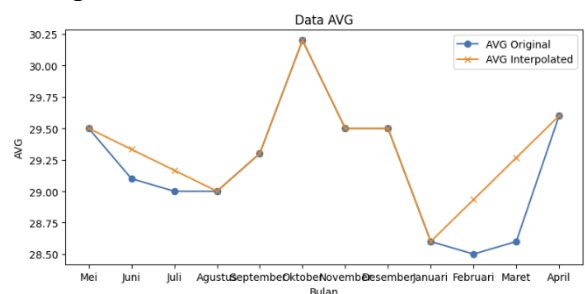
- b. Suhu Maksimum Tertinggi (HMAX)

Interpolasi HMAX cukup akurat untuk sebagian besar bulan. Namun, ada perbedaan signifikan pada bulan Juli, di mana data asli menunjukkan puncak yang tidak terdeteksi oleh interpolasi. Ini menunjukkan bahwa metode interpolasi linier mungkin kurang efektif dalam menangkap nilai ekstrem.



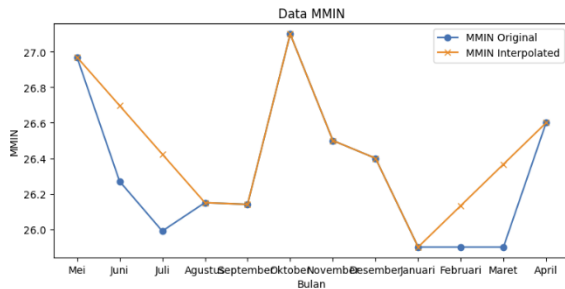
- c. Suhu Rata-rata (AVG)

Hasil interpolasi untuk AVG sangat mendekati data asli. Ini mengindikasikan bahwa metode interpolasi linier sangat efektif dalam memperkirakan suhu rata-rata.



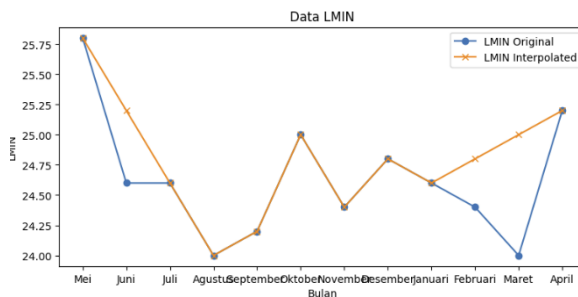
#### d. Suhu Minimum Rata-rata (MMIN)

Interpolasi MMIN menunjukkan hasil yang cukup akurat, meskipun terdapat sedikit perbedaan pada bulan Juni dan Juli. Pola umum dari data asli berhasil ditangkap oleh interpolasi.



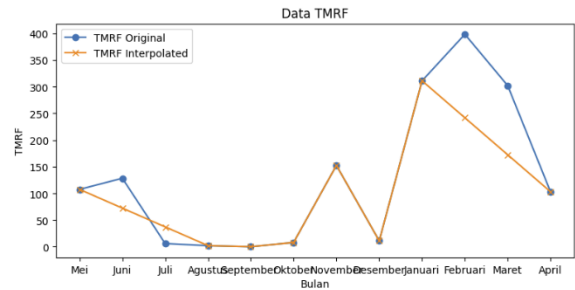
#### e. Suhu Minimum Terendah (LMIN)

Hasil interpolasi LMIN cukup baik, namun ada perbedaan yang signifikan pada bulan Juli. Interpolasi tidak menangkap penurunan suhu yang terjadi pada data asli di bulan tersebut.



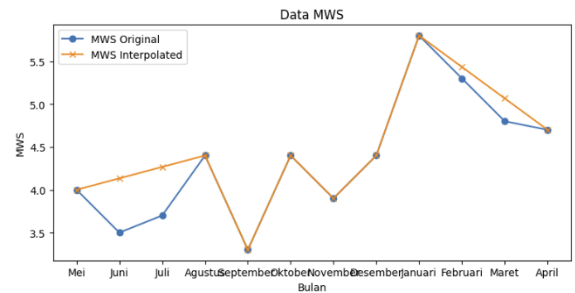
#### f. Total Curah Hujan (TMRF)

Interpolasi TMRF menunjukkan perbedaan yang paling signifikan antara data asli dan hasil interpolasi. Terlihat bahwa interpolasi linier kesulitan dalam memperkirakan pola curah hujan yang sangat bervariasi, terutama pada bulan Februari dan Maret di mana terjadi peningkatan curah hujan yang drastis pada data asli.



#### g. Kecepatan Angin Rata-rata (MWS)

Hasil interpolasi MWS cukup mendekati data asli, meskipun terdapat beberapa perbedaan, terutama pada bulan Juli dan Agustus. Pola umum kecepatan angin berhasil ditangkap oleh interpolasi.



### Analisis Kesalahan

Untuk mengevaluasi akurasi metode interpolasi, kami menghitung Mean Absolute Percentage Error (MAPE) dan Root Mean Square Percentage Error (RMSPE) untuk setiap variabel. Hasil ini dapat dilihat pada bagian bawah Tabel 2. Berikut adalah ringkasannya:

Variabel	MAPE (%)	RMSPE (%)
AVG	0.57	1.14
MMA	0.81	1.75
MMIN	0.44	0.85
TMRF	0.5	0.89
MWS	3.46	7.07

Tabel 2. Perhitungan MAPE dan RMSPE

Dari hasil ini, kita dapat menyimpulkan bahwa:

1. Suhu minimum rata-rata (MMIN) memiliki tingkat akurasi tertinggi dengan MAPE 0.44% dan RMSPE 0.85%.
2. Kecepatan angin rata-rata (MWS) memiliki tingkat akurasi terendah dengan MAPE 3.46% dan RMSPE 7.07%.
3. Secara umum, interpolasi untuk parameter suhu (AVG, MMAX, MMIN) memiliki akurasi yang lebih tinggi dibandingkan dengan parameter curah hujan (TMRF) dan kecepatan angin (MWS).

### Diskusi

Berdasarkan hasil yang ditampilkan di Image 1 dan Image 2, kita dapat menyimpulkan beberapa poin penting:

1. Metode interpolasi linier cukup efektif untuk memperkirakan data yang hilang, terutama untuk parameter suhu (AVG, MMAX, MMIN). Ini terlihat dari nilai MAPE dan RMSPE yang rendah untuk parameter-parameter tersebut.
2. Parameter curah hujan (TMRF) dan kecepatan angin (MWS) menunjukkan variabilitas yang lebih tinggi dan lebih sulit diprediksi menggunakan interpolasi linier. Ini terlihat dari grafik TMRF yang menunjukkan perbedaan signifikan antara data asli dan hasil interpolasi, serta nilai MAPE dan RMSPE yang lebih tinggi untuk MWS.

3. Metode ini kurang akurat dalam memperkirakan nilai ekstrem atau perubahan mendadak dalam data, seperti yang terlihat pada grafik HMAX untuk bulan Juli dan TMRF untuk bulan Februari dan Maret.
4. Meskipun interpolasi linier memberikan hasil yang cukup baik untuk sebagian besar parameter, metode ini mungkin tidak selalu menangkap pola musiman atau tren jangka panjang yang kompleks.

### KESIMPULAN

Penelitian ini bertujuan untuk menganalisis data curah hujan yang hilang menggunakan metode interpolasi linier. Dataset yang digunakan data curah hujan daerah Jakarta pada bulan Mei 2023 hingga April 2024 dengan beberapa parameter (Suhu rata-rata (AG), Suhu maksimum rata-rata (MMAX), Suhu maksimum tertinggi (HMAX), Suhu minimum rata-rata (MMIN), Suhu minimum terendah (LMIN), Total curah hujan (TMRF), Kecepatan Angin rata-rata (MWS)). Penelitian ini dilakukan dengan menggunakan bahasa pemrograman python dan jupyter notebook untuk mencari dataset yang kosong pada bulan tertentu. Hasil penelitian menunjukkan bahwa akurasi pencarian dataset yang hilang menggunakan metode interpolasi linier lebih baik dari pada metode interpolasi lainnya. Hal ini dibuktikan dari hasil Mean Square Error (MSE) metode interpolasi linier yang lebih rendah pada metode lainnya.

## DAFTAR PUSTAKA

[1] M. Rizky Ismail, A. Zain, F. Dewantoro, D. Pratiwi, dan T. Sipil, "Perhitungan Data Curah Hujan yang Hilang dengan Menggunakan Metode Interpolasi Linier." [Daring]. Tersedia pada:

<http://jim.teknokrat.ac.id/index.php/tekniksipilJurnalTeknikSipil>

[2] A. Marbun dan D. Nofriansyah, "ANALISA DATA MINING UNTUK MENGESTIMASI POTENSI CURAH HUJAN DENGAN MENGGUNAKAN METODE REGRESI LINEAR BERGANDA STMIK Triguna Dharma \*\* Program Studi Sistem Informasi, STMIK Triguna Dharma \*\*\* Program Studi Sistem Informasi, STMIK Triguna Dharma Keyword: Curah Hujan Data Mining Regresi Linear Berganda," *Jurnal CyberTech*, vol. 4, no. 2, 2021, [Daring]. Tersedia pada: <https://ojs.trigunadharma.ac.id/>