

- Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
    - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
    - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
  2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

# **IMPUTASI *MISSING DATA* CURAH HUJAN DI STASIUN BMKG JAWA TIMUR BERBASIS *MACHINE LEARNING* MENGUNAKAN DATA SATELIT**

**LILIS SRIWAHYUNI**



**PROGRAM STUDI MATEMATIKA TERAPAN  
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFORMATIKA  
INSTITUT PERTANIAN BOGOR  
BOGOR  
2025**



## @Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



## PERNYATAAN MENGENAI TESIS DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa tesis dengan judul “Imputasi *Missing Data* Curah Hujan di Stasiun BMKG Jawa Timur Berbasis *Machine Learning* Menggunakan Data Satelit” adalah karya saya dengan arahan dari dosen pembimbing dan belum diajukan dalam bentuk apa pun kepada perguruan tinggi mana pun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka di bagian akhir tesis ini.

Dengan ini saya melimpahkan hak cipta dari karya tulis saya kepada Institut Pertanian Bogor.

Bogor, 16 Juni 2025

Lilis Sriwahyuni  
G5501231011



## RINGKASAN

LILIS SRIWAHYUNI. Imputasi *Missing Data* Curah Hujan di Stasiun BMKG Jawa Timur Berbasis *Machine Learning* Menggunakan Data Satelit. Dibimbing oleh SRI NURDIATI dan ENDAR HASAFAH NUGRAHANI.

Estimasi dan pemantauan pola curah hujan penting untuk pemodelan sistem hidrologi serta untuk perencanaan dan pengelolaan sumber daya air di berbagai bidang masyarakat seperti dalam menghadapi bencana alam kekeringan maupun banjir. Data curah hujan diperoleh melalui alat pengukur hujan. Alat pengukur hujan mungkin tidak dapat mengidentifikasi curah hujan secara spasial di area yang luas atau di daerah dengan topografi yang kompleks. Kesalahan teknis dan kesalahan manusia juga sering kali mempengaruhi proses pencatatan data sehingga menyebabkan data menjadi tidak lengkap yang disebut dengan *missing data*.

Kebutuhan akan data amatan dari stasiun menjadi sangat penting dalam analisis data klimatologi, karena langsung diambil dari lokasi dan merupakan data aktual. Data stasiun di Indonesia banyak yang tidak lengkap sehingga analisis data curah hujan menggunakan data stasiun menjadi bermasalah. Beberapa peneliti menggunakan data satelit yang lebih lengkap untuk melakukan penelitian. Data satelit yang notabene bangkitan suatu model memiliki kesalahan sehingga hasil analisis juga mengandung kesalahan. Analisis menggunakan data stasiun lebih baik dibandingkan menggunakan data satelit. Salah satu cara untuk mengatasi data stasiun yang tidak lengkap adalah imputasi *missing data*. Imputasi adalah suatu proses untuk mengisi nilai yang kosong dengan nilai perkiraan yang dihitung secara tepat.

Penelitian ini memanfaatkan salah satu cabang *artificial intelligence* yaitu *machine learning* untuk melakukan imputasi *missing data* berdasarkan pemetaan data satelit pada data stasiun yang ada di Jawa Timur. Tujuan dari penelitian ini adalah menentukan metode *machine learning* dan data satelit dengan performa terbaik berdasarkan nilai *Mean Absolute Error* (MAE) paling rendah untuk imputasi *missing data* curah hujan di stasiun BMKG Jawa Timur.

Penelitian ini secara keseluruhan melalui empat tahapan yaitu tahap persiapan meliputi pengumpulan dan praproses data. Data curah hujan yang digunakan diperoleh dari stasiun BMKG Jawa Timur dan data satelit (ERA5, ERA5 Land, CMORPH CRT, CMORPH BLD, dan CHIRPS). Tahap kedua yaitu konstruksi model. Metode yang digunakan dalam penelitian ini yaitu *Multiple Linear Regression* (MLR), *Multiple Layer Perceptron* (MLP), *Support Vector Regression* (SVR), dan *Convolutional Neural Network* (CNN). Tahap ketiga yaitu pelatihan dan pengujian model. Model dilatih atas dasar pembagian data sebagai data pelatihan dan pengujian dengan rasio sebesar 95:5%, 90:10%, dan 80:20%. Melalui proses konstruksi model dihasilkan 20 kombinasi model yang kemudian dipilih model dengan nilai MAE paling rendah sebagai model dengan performa terbaik.

Berdasarkan metode *machine learning*, MLP menunjukkan performa yang baik pada data pelatihan maupun data pengujian, ditunjukkan oleh nilai MAE yang relatif rendah dibandingkan dengan metode lainnya. Pada data satelit, CMORPH CRT menghasilkan nilai MAE terendah pada data pelatihan, namun mengalami kenaikan MAE yang signifikan pada data pengujian, mengindikasikan adanya



*overfitting*. Oleh karena itu, CMORPH CRT tidak diusulkan dalam penelitian ini. Sementara itu, data ERA5 menunjukkan nilai MAE yang cukup rendah dan stabil pada data pelatihan dan pengujian, sehingga lebih direkomendasikan untuk digunakan. Secara keseluruhan, hasil penelitian ini menunjukkan bahwa kombinasi model MLP dengan data satelit ERA5 memberikan performa terbaik berdasarkan nilai MAE terendah, sehingga digunakan untuk imputasi data curah hujan yang hilang pada data stasiun BMKG di Jawa Timur.

Kata Kunci: Data satelit, data stasiun, imputasi *missing data*, *machine learning*, *mean absolute error*

- Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
    - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
    - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
  2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



## SUMMARY

LILIS SRIWAHYUNI. Imputation Missing Daily Rainfall Data in Station BMKG East Java Based on Machine Learning Using Satellite Data. Supervised by SRI NURDIATI and ENDAR HASAFAH NUGRAHANI.

The estimation and monitoring of rainfall patterns are crucial for hydrological system modeling, as well as for the planning and management of water resources across various societal sectors, such as community mitigation efforts in dealing with natural disasters like droughts and floods. Rainfall data is obtained through rain gauges. However, rain gauges may not accurately capture rainfall spatially across large areas or in regions with complex topography. Technical errors and human errors also frequently affect the data recording process, leading to incomplete data, commonly referred to as missing data.

Observational data from weather stations are essential in climatological data analysis, as they are collected directly from specific locations and represent actual data. Many station datasets in Indonesia are incomplete, which makes rainfall data analysis using station data problematic. Some researchers use more complete satellite data for their studies. Satellite data, which is typically generated by a model, contains errors, meaning that the analysis results also carry errors. Analysis using station data is considered better than using satellite data. One way to address the issue of incomplete station data is through missing data imputation. Imputation is a process of filling in missing values with estimated values calculated accurately.

This study utilizes a branch of artificial intelligence, namely machine learning, to perform missing data imputation based on satellite data mapping for station data in East Java. The aim of this study is to identify the machine learning method and satellite dataset with the best performance, determined by the lowest Mean Absolute Error (MAE), for imputing missing rainfall data at BMKG stations in East Java.

This study consists of four main stages. The first stage is preparation, which includes collecting and preprocessing data. The rainfall data used in this study were obtained from BMKG stations in East Java and from satellite datasets (ERA5, ERA5 Land, CMORPH CRT, CMORPH BLD, and CHIRPS). The second stage is model construction. The methods employed in this study are Multiple Linear Regression (MLR), Multilayer Perceptron (MLP), Support Vector Regression (SVR), and Convolutional Neural Network (CNN). The third stage involves model training and testing. The models were trained using data splits with training and testing ratios of 95:5%, 90:10%, and 80:20%. Through the model construction process, a total of 20 model combinations were produced, and the model with the lowest MAE was selected as the best-performing model.

Based on the machine learning methods, MLP demonstrated good performance on both the training and testing data, as indicated by its relatively low MAE compared to other methods. Among the satellite datasets, CMORPH CRT achieved the lowest MAE on the training data but showed a significant increase in MAE on the testing data, indicating overfitting. Therefore, CMORPH CRT is not recommended in this study. Meanwhile, the ERA5 dataset produced relatively low and stable MAE values across both training and testing data, making it more suitable for use. Overall, the results of this study show that the combination of the



MLP model and ERA5 satellite data provides the best performance, based on the lowest MAE value, and is thus used for imputing missing rainfall data at BMKG stations in East Java.

**Keywords:** Imputation missing data, machine learning, mean absolute error, satellite data, station data.

- Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
    - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
    - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
  2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



- Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
    - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
    - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
  2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

© Hak Cipta milik IPB, tahun 2025  
Hak Cipta dilindungi Undang-Undang

*Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan atau menyebutkan sumbernya. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik, atau tinjauan suatu masalah, dan pengutipan tersebut tidak merugikan kepentingan IPB.*

*Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apa pun tanpa izin IPB.*

# **IMPUTASI *MISSING DATA* CURAH HUJAN DI STASIUN BMKG JAWA TIMUR BERBASIS *MACHINE LEARNING* MENGUNAKAN DATA SATELIT**

**LILIS SRIWAHYUNI**

Tesis  
sebagai salah satu syarat untuk memperoleh gelar  
Magister Matematika pada  
Program Studi Matematika Terapan

**PROGRAM STUDI MATEMATIKA TERAPAN  
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFORMATIKA  
INSTITUT PERTANIAN BOGOR  
BOGOR  
2025**



**@Hak cipta milik IPB University**

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
  - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

Tim Penguji pada Ujian Tesis:

- 1 Elis Khatizah, S.Si., M.Si., Ph.D.
- 2

Judul Tesis : Imputasi *Missing Data* Curah Hujan di Stasiun BMKG Jawa Timur Berbasis *Machine Learning* Menggunakan Data Satelit  
Nama : Lilis Sriwahyuni  
NIM : G5501231011

Disetujui oleh

Pembimbing 1:  
Prof. Dr. Ir. Sri Nurdianti, M.Sc.



Pembimbing 2:  
Prof. Dr. Ir. Endar Hasafah Nugrahani, M.S.



Diketahui oleh

Ketua Program Studi:  
Prof. Dr. Drs. Jaharuddin, M.S.  
NIP 19651102 199302 1 001



Dekan Sekolah Sains Data, Matematika, dan Informatika  
Prof. Dr. Ir. Agus Buono, M.Si., M.Kom.  
NIP 19660702 199301 1 001



Tanggal Ujian: 27 Mei 2025

Tanggal Lulus: 17 JUN 2025



## PRAKATA

Puji syukur ke hadirat Allah SWT yang telah melimpahkan rahmat dan karunia-Nya, sehingga penulis dapat menyelesaikan tesis yang berjudul “Imputasi *Missing Data* Curah Hujan di Stasiun BMKG Jawa Timur Berbasis *Machine Learning* Menggunakan Data Satelit”. Shalawat dan salam senantiasa penulis haturkan kepada Nabi Muhammad SAW, seorang rasul akhir zaman yang menuntun ke jalan lurus dan penyeru kepada agama yang benar.

Penyusunan tesis ini tentunya tidak terlepas dari dukungan, bimbingan, dan bantuan berbagai pihak. Oleh karena itu, pada kesempatan ini penulis ingin menyampaikan penghargaan dan ucapan terima kasih yang sebesar-besarnya kepada:

1. Prof. Dr. Ir. Sri Nurdianti, M.Sc. dan Prof. Dr. Ir. Endar Hasafah Nugrahani, M.S. selaku komisi pembimbing, yang telah memberikan ilmu, masukan, arahan, dan motivasi selama proses penyusunan tesis ini.
2. Dr. Ir. I Gusti Putu Purnaba DEA. selaku pembahas dalam seminar kolokium dan, Elis Khatizah, S.Si., M.Si., Ph.D. selaku penguji luar komisi dalam ujian tesis yang telah memberikan banyak kritik, saran, dan wawasan dalam penulisan karya ilmiah ini.
3. Prof. Dr. Drs. Jaharuddin, M.S. selaku moderator seminar kolokium dan Prof. Dr. Dra. R. Iis Arifiantini M.Si. selaku dosen seminar hasil.
4. M. Khairun Najib, S.Si., M.Mat. selaku asisten dosen pembimbing, yang telah membantu dalam banyak hal hingga tesis ini selesai disusun.
5. Ayahanda dan Ibunda terkasih H. Danum, S.Pd. dan Hj. Sriwasih serta kakak-kakak tercinta Muhammad Wahyudi, S.T., Zikri Hariyadi S.Si., dan Sirlaily Fitriana S.Pd., yang selalu memberikan kasih sayang, doa, dan dukungan baik dari segi materi, moral maupun spiritual.
6. Seluruh staf pengajar dan akademik Klaster Matematika SSMI IPB University yang telah membantu penulis selama menempuh pendidikan di IPB.
7. Teman-teman Program Studi S2 Matematika Terapan angkatan 60, selaku teman diskusi dan seperjuangan.
8. Semua pihak lain yang tidak dapat disebutkan satu per satu yang telah membantu penulisan karya ilmiah ini.

Semoga Allah *Subhanahu Wa Ta’ala* melimpahkan rahmat, taufik, dan hidayah serta inayah-Nya kepada kita semua, Amiin.

Semoga karya ilmiah ini dapat bermanfaat bagi pihak yang membutuhkan dan bagi kemajuan ilmu pengetahuan khususnya dalam bidang matematika dan terapannya.

Bogor, 16 Juni 2025

*Lilis Sriwahyuni*

## DAFTAR ISI

|                                                                     |     |
|---------------------------------------------------------------------|-----|
| DAFTAR TABEL                                                        | xiv |
| DAFTAR GAMBAR                                                       | xiv |
| I PENDAHULUAN                                                       | 1   |
| 1.1 Latar Belakang                                                  | 1   |
| 1.2 Rumusan Masalah                                                 | 3   |
| 1.3 Tujuan Penelitian                                               | 3   |
| 1.4 Manfaat Penelitian                                              | 3   |
| II TINJAUAN PUSTAKA                                                 | 4   |
| 2.1 Missing Data                                                    | 4   |
| 2.2 Pendekatan Imputasi <i>Missing Data</i>                         | 4   |
| 2.3 Metode-metode Imputasi <i>Missing Data</i>                      | 5   |
| 2.4 <i>Loss Function</i>                                            | 9   |
| 2.5 <i>Optimizer</i>                                                | 10  |
| 2.6 Metrik Evaluasi                                                 | 11  |
| III METODE                                                          | 12  |
| 3.1 Sumber dan Jenis Data                                           | 12  |
| 3.2 Tahapan Penelitian                                              | 14  |
| IV HASIL DAN PEMBAHASAN                                             | 17  |
| 4.1 Praproses Data                                                  | 17  |
| 4.2 Konstruksi Model                                                | 19  |
| 4.3 Pelatihan dan Pengujian Model                                   | 25  |
| 4.4 Analisis Hasil                                                  | 29  |
| 4.5 Performa <i>Loss</i> Berdasarkan Metode <i>Machine Learning</i> | 30  |
| 4.6 Performa <i>Loss</i> Berdasarkan Jenis Data Satelit             | 31  |
| V SIMPULAN DAN SARAN                                                | 33  |
| 5.1 Simpulan                                                        | 33  |
| 5.2 Saran                                                           | 33  |
| DAFTAR PUSTAKA                                                      | 34  |
| LAMPIRAN                                                            | 38  |
| RIWAYAT HIDUP                                                       | 39  |



## DAFTAR TABEL

|    |                                                     |    |
|----|-----------------------------------------------------|----|
| 1  | Rumus pembaruan <i>optimizer</i>                    | 11 |
| 2  | Spesifikasi data stasiun                            | 12 |
| 3  | Spesifikasi data satelit                            | 13 |
| 4  | Ukuran matriks data satelit setelah praproses       | 17 |
| 5  | Percobaan <i>optimizer</i> dan <i>learning rate</i> | 19 |
| 6  | Percobaan parameter <i>cost</i> dan <i>gamma</i>    | 25 |
| 7  | <i>Training model</i> menggunakan data ERA5         | 25 |
| 8  | <i>Training model</i> menggunakan CNN               | 26 |
| 9  | <i>Training model</i> menggunakan MLR               | 26 |
| 10 | <i>Training model</i> menggunakan MLP               | 27 |
| 11 | <i>Training model</i> menggunakan SVR               | 28 |

## DAFTAR GAMBAR

|    |                                                                 |    |
|----|-----------------------------------------------------------------|----|
| 1  | Prosedur CNN pada dua dimensi                                   | 7  |
| 2  | Lokasi stasiun cuaca observasi Jawa Timur                       | 13 |
| 3  | Diagram alir penelitian                                         | 16 |
| 4  | Cakupan wilayah pengamatan data satelit                         | 18 |
| 5  | Arsitektur model CNN                                            | 21 |
| 6  | Arsitektur model MLR                                            | 22 |
| 7  | Arsitektur model MLP                                            | 23 |
| 8  | Imputasi <i>missing data</i> pada beberapa stasiun              | 29 |
| 9  | Performa <i>loss</i> berdasarkan metode <i>machine learning</i> | 30 |
| 10 | Performa <i>loss</i> berdasarkan jenis data satelit             | 31 |