

PERBANDINGAN METODE PENDUGAAN DATA HILANG PADA DATA IKLIM BALIKPAPAN

AMALIA SAFIRA WIDYAWATI



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

- 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 SKRIPSI DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa skripsi dengan judul “Perbandingan Metode Pendugaan Data Hilang pada Data Iklim Balikpapan” 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 skripsi ini.

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

Bogor, September 2025

Amalia Safira Widyawati
G1401211088

- 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.

ABSTRAK

AMALIA SAFIRA WIDYAWATI. Perbandingan Metode Pendugaan Data Hilang pada Data Iklim Balikpapan. Dibimbing oleh ANWAR FITRIANTO dan PIKA SILVIANI.

Data iklim memiliki peran penting dalam berbagai bidang kehidupan. Namun, sering kali ditemukan data hilang yang dapat mengganggu proses pengolahan data dan menurunkan kualitas analisis. Oleh karena itu, diperlukan metode penanganan yang tepat agar hasil analisis tetap valid. Penelitian ini bertujuan membandingkan performa beberapa metode imputasi untuk data hilang multivariat berdasarkan identifikasi pola hilang data aktual, serta menentukan metode yang paling sesuai dengan mekanisme data hilangnya. Metode yang digunakan meliputi *mean* per bulan, *missRanger*, *k-Nearest Neighbor* (*k*-NN), dan *Iterative Robust-Model Imputation* (IRMI). Informasi data hilang diperoleh dari Data *Global Surface Summary of the Day* (GSOD), dengan proporsi data hilang sebesar 11,4%, lalu pola data hilang tersebut diterapkan pada Data NasaPower yang relatif lengkap untuk evaluasi hasil imputasi. Hasil penelitian menunjukkan bahwa IRMI kurang mampu menangani kondisi data hilang ekstrem. Sebaliknya, *k*-NN, *missRanger* dan *mean* per bulan memberikan hasil yang lebih baik pada kondisi ekstrem maupun non-ekstrem. Dari keempat metode, *mean* per bulan terbukti paling unggul karena mampu mengatasi data hilang sekaligus mempertahankan struktur multivariat.

Kata kunci: data hilang multivariat, data iklim, imputasi, *k*-NN, *mean* per bulan

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan 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 mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

ABSTRACT

AMALIA SAFIRA WIDYAWATI. Comparison of Missing Data Estimation Methods in Balikpapan Climate Data. Supervised by ANWAR FITRIANTO and PIKA SILVIANTI.

Climate data has an important role in various areas of life. However, it is often found that missing values can interfere with the data processing process and reduce the quality of the analysis. Therefore, the right handling method is needed so that the analysis results remain valid. This study aims to compare the performance of several imputation methods for multivariate missing data based on the identification of actual missing data patterns, as well as determine the method that best suits the missing data mechanism. The methods used included *mean* per month, missRanger, k-Nearest Neighbor (k-NN), and Iterative Robust-Model Imputation (IRMI). The missing data information was obtained from the Global Surface Summary of the Day (GSOD) Data, with a proportion of missing data of 11.4%, then the missing data pattern was applied to the relatively complete NasaPower Data for the evaluation of imputation results. The results of the study show that IRMI is less able to handle extreme missing data conditions. In contrast, k-NN, missRanger and mean per month gave better results in both extreme and non-extreme conditions. Of the four methods, the mean per month proved to be the most superior because it was able to overcome missing data while maintaining a multivariate structure.

Keywords: average per month, climate data, imputation, k-NN, multivariate lost data



@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.

PERBANDINGAN METODE PENDUGAAN DATA HILANG PADA DATA IKLIM BALIKPAPAN

AMALIA SAFIRA WIDYAWATI

Skripsi
sebagai salah satu syarat untuk memperoleh gelar
Sarjana pada
Program Studi Statistika dan Sains Data

**PROGRAM STUDI SARJANA STATISTIKA DAN SAINS DATA
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.

Judul Skripsi : Perbandingan Metode Pendugaan Data Hilang pada Data Iklim
Balikpapan

Nama : Amalia Safira Widyawati

NIM : G1401211088

Disetujui oleh

Pembimbing 1:

Dr. Anwar Fitrianto, S.Si., M.Sc.



Pembimbing 2:

Pika Silvianti, S.Si., M.Si.



Diketahui oleh

Ketua Program Studi:

Dr. Bagus Sartono, S.Si., M.Si.

NIP 197804112005011002



Tanggal Ujian:
13 Agustus 2025

Tanggal Lulus:



- 

PRAKATA

Puji dan syukur penulis panjatkan kepada Allah subhanaahu wa ta'ala atas segala karunia-Nya sehingga karya ilmiah ini berhasil diselesaikan. Tema yang dipilih dalam penelitian yang dilaksanakan sejak bulan Maret 2025 sampai bulan Juli 2025 ini ialah data iklim, dengan judul “Perbandingan Metode Pendugaan Data Hilang pada Data Iklim Balikpapan”.

Terima kasih penulis ucapkan kepada semua pihak yang telah berperan aktif serta membantu penulis dalam penyusunan karya ilmiah ini, di antaranya :

1. Dr. Anwar Fitrianto, S.Si., M.Sc. dan Pika Silvianti, S.Si., M.Si., selaku komisi pembimbing yang telah banyak memberi saran, arahan, waktu dan tenaga dalam proses penulisan karya ilmiah ini.
2. Bapak Dr. Ir. Budi Susetyo, M.S, selaku penguji, Ibu Rahma Anisa, S.Stat., M.Si dan Bapak Dr. Bagus Sartono, S.Si., M.Si selaku moderator kolokium dan seminar hasil yang telah memberikan saran dan masukan untuk perkembangan karya ilmiah ini.
3. Bapak Gatot Subroto, Ibu Susilo Widayati, serta kakak-kaka saya dan seluruh keluarga atas segala doa dan dukungannya untuk penulis dalam menyelesaikan karya ilmiah ini.
4. Seluruh dosen, staff dan TU Departemen Statistika FMIPA IPB yang telah membantu penulis selama jenjang perkuliahan hingga dapat menyelesaikan pendidikan dan lancarnya proses administrasi hingga ke tahap akhir.
5. Dindana, Azra, Nazuwa, Hanifa, Pingkan, Rani, Zafira, Ghina, dan Muti yang selalu memberikan dukungan, doa, serta apresiasi untuk semua tahapan yang pernah dilalui.
6. Sobat Alhur, warga Jonggol, dan teman-teman Gamavectra yang telah memberikan dukungan dan memberikan kebersamaan selama masa-masa kuliah dari semester awal hingga memasuki semester akhir ini.
7. Semua pihak yang tidak mungkin disebutkan satu per satu yang membantu, memberikan saran, arahan serta dukungan untuk penulis dalam menyelesaikan karya ilmiah sebagai tugas akhir ini.
8. Diri saya sendiri, yang masih berusaha bertahan meski sering ingin menyerah, sambil terus belajar agar bisa tumbuh dan pulih.

Semoga karya ilmiah ini bermanfaat bagi pihak yang membutuhkan dan bagi kemajuan ilmu pengetahuan.

Bogor, September 2025

Amalia Safira Widyawati



@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.

DAFTAR ISI

DAFTAR TABEL	xiii
DAFTAR GAMBAR	xiii
LAMPIRAN	xiii
I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Tujuan	2
II TINJAUAN PUSTAKA	3
2.1 Data Iklim	3
2.2 <i>k-Nearest Neighbor Imputation</i>	3
2.3 <i>missRanger</i>	4
2.4 IRMI	4
2.5 Mekanisme Data Hilang	4
2.6 Missing Completely At Random (MCAR)	5
2.6.1 Pengujian MCAR	5
2.6.2 Uji Non-Parametrik untuk MCAR	6
2.7 <i>Missing At Random (MAR)</i>	7
2.8 <i>Missing Not At Random (MNAR)</i>	7
III METODE	9
3.1 Data	9
3.2 Prosedur Analisis	9
IV HASIL DAN PEMBAHASAN	13
4.1 Eksplorasi Data <i>Global Summary of The Day (GSOD)</i>	13
4.1.1 Statistika Deskriptif	13
4.1.2 Visualisasi Setiap Peubah	13
4.1.3 Korelasi Antar Peubah	15
4.1.4 Visualisasi Data Hilang pada Data Iklim Balikpapan	16
4.2 Identifikasi Pola Data Hilang pada Data GSOD	17
4.2.1 MCAR	17
4.2.2 MAR	17
4.2.3 MNAR	18
4.3 Hasil Pendugaan Data Hilang pada Data NasaPower	19
4.4 Evaluasi Hasil Pendugaan	19
4.4.1 Evaluasi Numerik	20
4.4.2 Evaluasi Visual	22
4.5 Penerapan pada Data GSOD	23
V SIMPULAN DAN SARAN	25
5.1 Simpulan	25

5.2	Saran	25
	DAFTAR PUSTAKA	26
	LAMPIRAN	29

@Hak cipta milik IPB University

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.

DAFTAR TABEL

1	Peubah yang digunakan	9
2	Statistik deskriptif	13
3	Hasil uji MCAR	17
4	Hasil model regresi logistik untuk identifikasi mekanisme MAR	18
5	Hasil pendugaan data hilang setiap metode	19
6	Nilai metrik evaluasi hasil imputasi	20
7	Nilai metrik evaluasi hasil imputasi tanpa pola hilang keseluruhan	21
8	Perbandingan Statistik Deskriptif Peubah Curah Hujan	23

DAFTAR GAMBAR

1	Diagram alir tahapan analisis data	12
2	Grafik sebaran data	14
3	Grafik tren peubah Prec (curah hujan)	14
4	Korelasi antar peubah	15
5	Peta data hilang	16
6	Pola data hilang	16
7	Evaluasi visual setiap metode	22

LAMPIRAN

1	Penentuan Nilai k pada metode k -NN	30
2	<i>Pair Plot</i> antar peubah Data NasaPower	31
3	Perbandingan Statistik Deskriptif Data GSOD Sebelum dan Sesudah Imputasi	32
4	Plot Kepekatan Curah Hujan Data GSOD	33