

STATISTICAL DOWNSCALING KONSENTRASI PM2.5 DI JAKARTA MENGGUNAKAN XGBOOST BERBASIS IMPUTASI AOD MAIAC

RIDWAN CAHYA ALFIANDI



**PROGRAM SARJANA ILMU KOMPUTER
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
BOGOR
2026**



@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 SKRIPSI DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa skripsi dengan judul “*Statistical Downscaling* Konsentrasi PM2.5 di Jakarta Menggunakan XGBoost Berbasis Imputasi AOD MAIAC” 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, Juli 2026

Ridwan Cahya Alfiandi
G6401221006



ABSTRAK

RIDWAN CAHYA ALFIANDI. *Statistical Downscaling* Konsentrasi PM2.5 di Jakarta Menggunakan XGBoost Berbasis Imputasi AOD MAIAC. Dibimbing oleh AHMAD RIDHA dan MUHAMMAD ASYHAR AGMALARO.

Konsentrasi PM2.5 tahunan di Jakarta jauh melampaui pedoman WHO, namun pemantauan berbasis stasiun darat masih terbatas. *Aerosol Optical Depth* (AOD) dari produk MAIAC berpotensi melengkapi keterbatasan tersebut tetapi mengalami proporsi *missing value* yang sangat tinggi akibat tutupan awan di wilayah tropis. Penelitian ini bertujuan membangun dan mengevaluasi model imputasi AOD MAIAC serta model *statistical downscaling* konsentrasi PM2.5 menggunakan XGBoost di Jakarta selama periode Januari 2022 hingga Desember 2025. Proporsi *missing value* AOD berkisar antara 80,29% hingga 84,39%. Model imputasi AOD menggunakan prediktor aerosol MERRA-2 dan meteorologi ERA5 menghasilkan R^2 pada data pengujian sebesar 0,9518 hingga 0,9807. Data AOD hasil imputasi selanjutnya digunakan sebagai prediktor utama pada model *statistical downscaling* PM2.5 yang dikembangkan dengan sepuluh perlakuan konfigurasi fitur mencakup representasi spasial, fitur temporal, dan reduksi dimensi PCA. Perlakuan terbaik menghasilkan R^2 sebesar 0,7459 pada SPKU DKI1 Bundaran HI, 0,6841 pada SPKU DKI2 Kelapa Gading, dan 0,4861 pada SPKU DKI3 Jagakarsa. Penambahan fitur *lag* dan *rolling mean* meningkatkan performa pada dua stasiun, sedangkan PCA secara konsisten menurunkan performa. XGBoost terbukti efektif untuk imputasi AOD namun performa *downscaling* PM2.5 bervariasi antarstasiun dan dipengaruhi oleh tingginya proporsi data hasil imputasi serta karakteristik.

Kata kunci: AOD MAIAC, imputasi, PM2.5, *statistical downscaling*, XGBoost

ABSTRACT

RIDWAN CAHYA ALFIANDI. *Statistical Downscaling* of PM2.5 Concentrations in Jakarta Using XGBoost Based on MAIAC AOD Imputation. Supervised by AHMAD RIDHA and MUHAMMAD ASYHAR AGMALARO.

Jakarta's annual PM2.5 concentration far exceeds the WHO guideline, yet ground-based monitoring remains limited. Aerosol Optical Depth (AOD) from the MAIAC product can potentially supplement this limitation but suffers from very high missing value proportions due to cloud cover in tropical regions. This study aimed to develop and evaluate MAIAC AOD imputation and PM2.5 statistical downscaling models using XGBoost in Jakarta from January 2022 to December 2025. Missing value proportions of AOD ranged from 80,29% to 84,39%. The AOD imputation model was built using MERRA-2 aerosol and ERA5 meteorological predictors, achieving test R^2 values of 0,9518 to 0,9807. The imputed AOD data were subsequently used as the primary predictor in the PM2.5

statistical downscaling model, which was developed with ten feature configuration treatments encompassing spatial representation, temporal features, and PCA dimensionality reduction. The best treatments yielded R^2 values of 0,7459 at DKI1 Bundaran HI, 0,6841 at DKI2 Kelapa Gading, and 0,4861 at DKI3 Jagakarsa. Lag and rolling mean features improved performance at two stations, while PCA consistently degraded performance. XGBoost proved effective for AOD imputation, but PM2.5 downscaling performance varied across stations, influenced by high imputed data proportions.

Keywords: AOD MAIAC, imputation, PM2.5, statistical downscaling, XGBoost

@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 mengumunkan 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 2026
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.

***STATISTICAL DOWNSCALING* KONSENTRASI PM_{2.5} DI JAKARTA MENGGUNAKAN XGBOOST BERBASIS IMPUTASI AOD MAIAC**

RIDWAN CAHYA ALFIANDI

Skripsi
sebagai salah satu syarat untuk memperoleh gelar
Sarjana pada
Program Studi Ilmu Komputer

**PROGRAM SARJANA ILMU KOMPUTER
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
BOGOR
2026**



@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 Skripsi:
Dr. Aziz Kustiyo, S.Si., M.Kom.



@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 : *Statistical Downscaling* Konsentrasi PM2.5 di Jakarta
Menggunakan XGBoost Berbasis Imputasi AOD MAIAC

Nama : Ridwan Cahya Alfiandi
NIM : G6401221006

@Hak cipta milik IPB University

Disetujui oleh

Pembimbing 1:
Ahmad Ridha, S.Kom., M.S.
19800507 200501 1 001



Pembimbing 2:
Muhammad Asyhar Agmalaro, S.Si., M.Kom.
19860331 201212 1 001



Diketahui oleh

Ketua Program Sarjana Ilmu Komputer:
Dr. Sony Hartono Wijaya, S.Kom., M.Kom.
19810809 200812 1 002



Tanggal Ujian:
23 Juli 2026

Tanggal Lulus:

PRAKATA

Puji dan syukur penulis panjatkan kepada Allah *subhanahu wa ta'ala* atas segala karunia-Nya sehingga karya ilmiah ini berhasil diselesaikan. Tema yang dipilih dalam penelitian yang dilaksanakan sejak bulan Oktober 2025 sampai bulan Juli 2026 ini ialah *statistical downscaling* konsentrasi PM2.5, dengan judul "*Statistical Downscaling* Konsentrasi PM2.5 di Jakarta Menggunakan XGBoost Berbasis Imputasi AOD MAIAC". Penulis menyampaikan terima kasih yang sebesar-besarnya kepada dosen pembimbing, Bapak Ahmad Ridha, S.Kom., M.S., dan Bapak Muhammad Asyhar Agmalaro, S.Si., M.Kom., atas bimbingan, arahan, dan dukungan yang diberikan selama proses penelitian ini.

Penulis mengakui bahwa penyelesaian karya ilmiah ini tidak dapat terwujud tanpa dukungan, bantuan, bimbingan, dan saran dari berbagai pihak selama proses penyusunan. Ucapan terima kasih yang mendalam juga penulis sampaikan kepada:

- a. Bapak Tjahyo dan Ibu Ida selaku kedua orang tua penulis, serta Kak Arief, atas doa, dukungan, dan motivasi yang senantiasa diberikan selama proses penelitian,
- b. Keluarga besar penulis yang turut memberikan dukungan hingga penulis sampai pada titik ini,
- c. Sahabat dan rekan seperjuangan penulis, yaitu Raihana, Zhafran Agus, Izza, Syira, Khansa, Haris, Tiffany, Habib, Wisnu, Maulana, Kak Tata, Pandu Arga, Bintang, Ilkomerz 59, dan teman-teman lainnya yang selalu membantu dan mendukung penulis selama penyusunan karya ilmiah ini,
- d. Teman-teman SMA penulis, yaitu Adham, Faiz, Rijal, Rendra, Nurul, Caca, Muflih, Kesa, Atha, dan Sera, yang senantiasa berkomunikasi dan memberikan dukungan moral selama proses penyusunan skripsi ini, dan
- e. Semua pihak terkait yang tidak dapat disebutkan satu per satu yang telah memberikan dukungan demi kelancaran penyusunan karya ilmiah ini.

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

Bogor, Juli 2026

Ridwan Cahya Alfiandi



DAFTAR ISI

DAFTAR TABEL	xi
DAFTAR GAMBAR	xi
DAFTAR LAMPIRAN	xii
PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Tujuan	2
1.4 Manfaat	2
1.5 Ruang Lingkup	3
TINJAUAN PUSTAKA	4
2.1 <i>Particulate Matter 2.5 (PM2.5)</i>	4
2.2 AOD MAIAC	4
2.3 <i>Extreme Gradient Boosting (XGBoost)</i>	5
2.4 <i>Hyperparameter Tuning</i> dengan Optuna	7
METODE	8
3.1 Data Penelitian	8
3.2 Tahapan Penelitian	9
3.3 Peralatan Penelitian	14
HASIL DAN PEMBAHASAN	15
4.1 Praproses dan Eksplorasi Data	15
4.2 Imputasi AOD MAIAC Menggunakan XGBoost	17
4.3 <i>Statistical Downscaling</i> PM2.5 Menggunakan XGBoost	25
SIMPULAN DAN SARAN	35
5.1 Simpulan	35
5.2 Saran	35
DAFTAR PUSTAKA	36
LAMPIRAN	38
RIWAYAT HIDUP	45



DAFTAR TABEL

1	Konversi nilai konsentrasi PM2.5 dan kategori Indeks Standar Pencemar Udara (ISPU)	4
2	Ringkasan karakteristik data penelitian yang dikumpulkan	8
3	<i>Longitude</i> dan <i>latitude</i> stasiun pengukuran di Jakarta	9
4	Rentang pencarian <i>hyperparameter</i> yang digunakan	11
5	Konfigurasi sepuluh perlakuan <i>statistical downscaling</i>	13
6	Ringkasan hasil pembersihan data PM2.5 per stasiun	15
7	Statistik deskriptif PM2.5 harian per stasiun setelah pembersihan	15
8	Proporsi <i>missing value</i> (%) AOD MAIAC per <i>grid</i> setiap stasiun	16
9	Variabel yang digunakan pada model imputasi AOD MAIAC	17
10	Performa model imputasi AOD pada data pengujian per SPKU	24
11	Variabel dasar yang digunakan pada model <i>statistical downscaling</i> PM2.5	25
12	Jumlah fitur pada setiap perlakuan <i>statistical downscaling</i> PM2.5	26
13	Ringkasan hasil PCA Perlakuan 9 dan Perlakuan 10 setiap SPKU	27
14	Performa sepuluh perlakuan model <i>statistical downscaling</i> PM2.5 pada data pengujian di SPKU DKI1 Bundaran HI	27
15	Performa sepuluh perlakuan model <i>statistical downscaling</i> PM2.5 pada data pengujian di SPKU DKI2 Kelapa Gading	29
16	Performa sepuluh perlakuan model <i>statistical downscaling</i> PM2.5 pada data pengujian di SPKU DKI3 Jagakarsa	31
17	Ringkasan perlakuan terbaik model <i>statistical downscaling</i> PM2.5 per stasiun	33
18	Perbandingan performa <i>statistical downscaling</i> PM2.5 dengan penelitian terdahulu	34

DAFTAR GAMBAR

1	Struktur algoritma XGBoost (Amjad <i>et al.</i> 2022)	6
2	Diagram alur tahapan penelitian	9
3	Ilustrasi konsep <i>statistical downscaling</i> PM2.5 menggunakan XGBoost	13
4	Ketersediaan data AOD per bulan pada <i>grid aod_11</i> di SPKU DKI1 Bundaran HI	16
5	<i>Scatter plot</i> nilai AOD estimasi terhadap aktual pada data pengujian di SPKU DKI1 Bundaran HI	18
6	<i>Feature importance</i> pada pemodelan imputasi AOD di SPKU DKI1 Bundaran HI	18
7	Perbandingan <i>time series</i> nilai AOD sebelum dan setelah imputasi pada <i>grid aod_11</i> di SPKU DKI1 Bundaran HI	19
8	<i>Scatter plot</i> nilai AOD estimasi terhadap aktual pada data pengujian di SPKU DKI2 Kelapa Gading	20
9	<i>Feature importance</i> pada pemodelan imputasi AOD di SPKU DKI2 Kelapa Gading	20

10	Perbandingan <i>time series</i> nilai AOD sebelum dan setelah imputasi pada <i>grid aod_11</i> di SPKU DKI2 Kelapa Gading	21
11	<i>Scatter plot</i> nilai AOD estimasi terhadap aktual pada data pengujian di SPKU DKI3 Jagakarsa	22
12	<i>Feature importance</i> pada pemodelan imputasi AOD di SPKU DKI3 Jagakarsa	22
13	Perbandingan <i>time series</i> nilai AOD sebelum dan setelah imputasi pada <i>grid aod_11</i> di SPKU DKI3 Jagakarsa	23
14	Perbandingan <i>scatter plot</i> estimasi terhadap aktual konsentrasi PM2.5 pada (a) model dasar dan (b) model terbaik, di SPKU DKI1 Bundaran HI	28
15	Perbandingan <i>feature importance</i> pada (a) model dasar dan (b) model terbaik, di SPKU DKI1 Bundaran HI	28
16	Perbandingan <i>scatter plot</i> estimasi terhadap aktual konsentrasi PM2.5 pada (a) model dasar dan (b) model terbaik, di SPKU DKI2 Kelapa Gading	30
17	Perbandingan <i>feature importance</i> pada (a) model dasar dan (b) model terbaik, di SPKU DKI2 Kelapa Gading	30
18	Perbandingan <i>scatter plot</i> estimasi terhadap aktual konsentrasi PM2.5 pada (a) model dasar dan (b) model terbaik, di SPKU DKI3 Jagakarsa	32
19	Perbandingan <i>feature importance</i> pada (a) model dasar dan (b) model terbaik, di SPKU DKI3 Jagakarsa	32

DAFTAR LAMPIRAN

1	Statistik deskriptif variabel MERRA-2, ERA5, dan NDVI di setiap SPKU setelah praproses data	39
2	<i>Feature importance</i> pada sepuluh perlakuan model <i>statistical downscaling</i> PM2.5 setiap SPKU	41