

PEMODELAN KONSENTRASI PM_{2.5} DI INDONESIA DENGAN *GEOGRAPHICALLY AND TEMPORALLY WEIGHTED RANDOM FOREST* DAN INTERPRETASI BERBASIS SHAP

SINTONG M.N PURBA



**PROGRAM STUDI SARJANA STATISTIKA DAN SAINS DATA
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, PENGGUNAAN AI SERTA PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa skripsi dengan judul “Pemodelan Konsentrasi PM2.5 di Indonesia dengan *Geographically and Temporally Weighted Random Forest* dan Interpretasi Berbasis SHAP” 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.

Dalam penyusunan karya ini, penulis menggunakan bantuan kecerdasan buatan berupa ChatGPT, Claude, dan Perplexity dalam pencarian referensi ilmiah, penyuntingan dan perbaikan redaksi, dan proses brainstorming dalam penulisan skripsi. Setelah menggunakan alat/layanan tersebut, penulis meninjau dan menyunting konten sesuai kebutuhan serta bertanggungjawab penuh atas isi karya tugas akhir ini.

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

Bogor, Juni 2026

Sintong M.N Purba
G1401221077



ABSTRAK

SINTONG M.N PURBA. Pemodelan Konsentrasi PM2.5 di Indonesia dengan *Geographically and Temporally Weighted Random Forest* dan Interpretasi Berbasis SHAP. Dibimbing oleh AKBAR RIZKI dan YENNI ANGRAINI.

Konsentrasi *particulate matter* berukuran kurang dari 2.5 mikrometer (PM2.5) merupakan polutan udara yang berdampak serius terhadap kesehatan manusia dan bersifat heterogen secara spasial dan temporal. Penelitian ini menerapkan model *Geographically and Temporally Weighted Random Forest* (GTWRF) untuk memodelkan konsentrasi PM2.5 di Indonesia menggunakan data *grid* beresolusi $0.1^\circ \times 0.1^\circ$ periode 2020–2024 dengan enam peubah prediktor meliputi suhu udara, tekanan udara, curah hujan, *leaf area index*, kecepatan angin, dan jumlah penduduk. Model GTWRF terbukti menghasilkan performa yang lebih unggul dengan nilai *root mean square error* (RMSE) sebesar 3.113 dan R^2 sebesar 0.903 dibandingkan dengan *random forest* global dengan nilai RMSE 5.306 dan R^2 0.703. Interpretasi model melalui nilai *Shapley Additive Explanations* (SHAP) yang diringkaskan menggunakan algoritme *clustering large applications* (CLARA) menghasilkan lima gerombol wilayah dengan karakteristik pengaruh peubah yang berbeda dan dinamis sepanjang periode pengamatan. Curah hujan dan kecepatan angin merupakan peubah paling dominan yang arah dan besar pengaruhnya sangat dikendalikan oleh siklus *La Niña* dan *El Niño*.

Kata kunci: GTWRF, PM2.5, *random forest*, SHAP, spasial-temporal



ABSTRACT

SINTONG M.N PURBA. Modeling PM2.5 Concentrations in Indonesia Using Geographically and Temporally Weighted Random Forest (GTWRF) and SHAP-Based Interpretation. Supervised by AKBAR RIZKI and YENNI ANGRAINI.

The concentration of particulate matter smaller than 2.5 micrometers (PM2.5) is an air pollutant that has a serious impact on human health and is heterogeneously distributed both spatially and temporally. This study applies the geographically and temporally weighted random forest (GTWRF) model to simulate PM2.5 concentrations in Indonesia using $0.1^\circ \times 0.1^\circ$ grid data for the period 2020–2024, with six predictors variables including air temperature, air pressure, rainfall, leaf area index, wind speed, and population. The GTWRF model proved to perform better than the global random forest, with a root mean square error (RMSE) value of 3.113 and an R^2 of 0.903, compared to an RMSE of 5.306 and an R^2 of 0.703. Model interpretation through Shapley additive explanations (SHAP) summarized using the algorithm *clustering large applications* (CLARA), resulted in five regional clusters characterized by different and dynamic influences of the variables throughout the observation period. Rainfall and wind speed were the most dominant variables, with their direction and magnitude of influence being highly controlled by the La Niña and El Niño cycles.

Keywords: GTWRF, PM2.5, random forest, SHAP, spatio-temporal.



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.

PEMODELAN KONSENTRASI PM_{2.5} DI INDONESIA DENGAN *GEOGRAPHICALLY AND TEMPORALLY WEIGHTED RANDOM FOREST* DAN INTERPRETASI BERBASIS SHAP

SINTONG M.N PURBA

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
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:

1. Rahma Anisa, S.Stat., M.Si., M.Act.Sc.



@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 : Pemodelan Konsentrasi PM2.5 di Indonesia dengan
Geographically and Temporally Weighted Random Forest dan
Interpretasi Berbasis SHAP

Nama : Sintong M.N Purba
NIM : G1401221077

@Hak cipta milik IPB University

Disetujui oleh

Pembimbing 1:

Akbar Rizki, S.Stat., M.Si.

Pembimbing 2:

Dr. Yenni Angraini, S.Si., M.Si.

Diketahui oleh

Kepala Program Studi Statistika dan Sains Data:

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

NIP. 197804112005011002

Tanggal Ujian:
3 Juni 2026

Tanggal Lulus:



PRAKATA

Puji dan syukur penulis panjatkan kepada Tuhan Yang Maha Esa atas segala karunia-Nya sehingga karya ilmiah ini berhasil diselesaikan. Tema yang dipilih dalam penelitian yang dilaksanakan sejak bulan Januari 2026 sampai bulan Mei 2026 ini ialah pemodelan spasial-temporal *machine learning*, dengan judul “Pemodelan Konsentrasi PM2.5 di Indonesia dengan *Geographically and Temporally Weighted Random Forest* dan Interpretasi Berbasis SHAP”.

Selama proses penulisan karya ilmiah ini, penulis tidak luput dari bimbingan, doa, dan dukungan dari berbagai pihak sehingga karya ilmiah ini dapat terselesaikan dengan baik. Oleh karena itu, penulis ingin menyampaikan rasa terima kasih kepada semua pihak terutama :

1. Bapak Hamonangan Purba dan Ibu Masta Simamora sebagai orang tua penulis, C.D Enjelina Purba, Elisabeth Purba, Sutra Ceria Purba, dan Berkat Purba sebagai saudara penulis, serta seluruh keluarga besar yang selalu memberikan dukungan, doa dan motivasi selama penyusunan karya ilmiah ini;
2. Ibu Akbar Rizki, S.Stat., M.Si., dan Ibu Dr. Yenni Angraini, S.Si., M.Si. selaku komisi pembimbing yang memberikan arahan dan bimbingan selama proses penulisan karya ilmiah ini;
3. Ibu Rahma Anisa, S.Stat., M.Si., M.Act.Sc. selaku penguji luar komisi, Bapak Dr. Agus Mohamad Soleh S.Si., M.T. selaku moderator kolokium, dan Bapak Dr. Farit Mochamad Afendi, S.Si., M.Si. selaku moderator seminar hasil yang telah memberikan saran dan masukan dalam penulisan karya ilmiah ini;
4. Seluruh dosen dan tenaga pendidik Program Studi Statistika dan Sains Data yang telah memberikan ilmu dan membantu selama proses perkuliahan penulis,
5. Nanda, Bintang, Jodi, Ria, Ghoni yang menjadi teman diskusi selama penulisan karya ilmiah ini,
6. Teman-teman Statistika angkatan 59 yang kebersamaan perkuliahan selama 4 tahun ini; dan
7. Seluruh pihak yang ikut mendukung penulis yang tidak dapat disebutkan satu per satu.

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

Bogor, Juni 2026

Sintong M.N Purba



DAFTAR ISI

DAFTAR TABEL	xi
DAFTAR GAMBAR	xi
PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Tujuan	2
TINJAUAN PUSTAKA	3
2.1 <i>Random Forest</i>	3
2.2 <i>Geographically and Temporally Weighted Random Forest</i> (GTWRF)	4
2.3 <i>Shapley additive explanations</i> (SHAP)	5
2.4 <i>Clustering large applications</i> (CLARA)	6
METODE	7
3.1 Data	7
3.2 Prosedur Analisis Data	7
HASIL DAN PEMBAHASAN	11
4.1 Eksplorasi Data	11
4.2 Pemodelan Regresi Linier	14
4.3 Pemodelan <i>Random Forest</i> Global	16
4.4 Pemodelan <i>Geographically and Temporally Weighted Random Forest</i> (GTWRF)	16
4.5 Perbandingan Model Terbaik pada <i>Random Forest</i> Global dan GTWRF	17
4.6 Interpretasi Keragaman Pengaruh Berdasarkan Nilai SHAP	18
SIMPULAN	33
5.1 Simpulan	33
DAFTAR PUSTAKA	34
RIWAYAT HIDUP	38



DAFTAR TABEL

1	Daftar peubah yang digunakan	7
2	Dugaan parameter regresi linier	15
3	Diagnostik sisaan model regresi linier	15
4	Nilai VIF masing-masing peubah prediktor	15
5	Perbandingan <i>bandwidth</i> terbaik berdasarkan nilai evaluasi R^2 dan RMSE	17
6	Perbandingan model terbaik pada random forest global dan GTWRF	17
7	Profil gerombol nilai SHAP model GTWRF tahun 2020	20
8	Profil gerombol nilai SHAP model GTWRF tahun 2021	22
9	Profil gerombol nilai SHAP model GTWRF tahun 2022	24
10	Profil gerombol nilai SHAP model GTWRF tahun 2023	27
11	Profil gerombol nilai SHAP model GTWRF tahun 2024	29

DAFTAR GAMBAR

1	Peta tematik konsentrasi PM2.5 di Indonesia tahun: (a) 2020, (b) 2021, (c) 2022, (d) 2023, dan (e) 2024	11
2	<i>Boxplot</i> peubah prediktor (a) X_1 , (b) X_2 , (c) X_3 , (d) X_4 , (e) X_5 , dan (f) X_6	12
3	Matriks korelasi antarpeubah yang digunakan	13
4	<i>Scatter plot</i> peubah respons terhadap peubah prediktor: (a) X_1 , (b) X_2 , (c) X_3 , (d) X_4 , (e) X_5 , dan (f) X_6	14
5	Evaluasi <i>bandwidth</i> melalui nilai (a) R^2 , dan (b) RMSE	16
6	Plot nilai <i>silhouette</i> tahun: (a) 2020, (b) 2021, (c) 2022, (d) 2023, dan (e) 2024	19
7	Peta tematik gerombol nilai SHAP model GTWRF tahun 2020	20
8	Peta tematik gerombol nilai SHAP model GTWRF tahun 2021	22
9	Peta tematik gerombol nilai SHAP model GTWRF tahun 2022	25
10	Peta tematik gerombol nilai SHAP model GTWRF tahun 2023	27
11	Peta tematik gerombol nilai SHAP model GTWRF tahun 2024	30



@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.