

# **PEMODELAN *STATISTICAL DOWNSCALING* DATA *SUNSHINE DURATION* MENGGUNAKAN RANDOM FOREST DI HIKMAH FARM PANGALENGAN**

**TIFFANY ANASTASIA JOCELYN**



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



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

Dengan ini saya menyatakan bahwa skripsi dengan judul “Pemodelan *Statistical Downscaling Data Sunshine Duration* Menggunakan Random forest di Hikmah Farm Pangalengan” 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

Tiffany Anastasia Jocelyn  
G6401221005

- 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

TIFFANY ANASTASIA JOCELYN. Pemodelan *Statistical Downscaling* Data *Sunshine Duration* Menggunakan Random Forest di Hikmah Farm Pangalengan. Dibimbing oleh MUHAMMAD ASYHAR AGMALARO dan AZIZ KUSTIYO.

Lama penyinaran matahari merupakan salah satu informasi agroklimat yang penting dalam mendukung pengelolaan budidaya dan pengambilan keputusan pertanian. Penelitian ini bertujuan membangun model klasifikasi kejadian *sunshine duration* lokal di Hikmah Farm Pangalengan menggunakan random forest dengan data atmosfer ERA5 sebagai variabel prediktor dan data *sunshine duration* Sinaubumi IPB sebagai variabel target biner. Pada penelitian ini, kelas 1 merepresentasikan kondisi terdapat lama penyinaran matahari pada jam yang diprediksi, sedangkan kelas 0 merepresentasikan kondisi tidak terdapat lama penyinaran matahari pada jam yang diprediksi. Data ERA5 terlebih dahulu diproses melalui interpolasi *Inverse Distance Weighted* (IDW) untuk meningkatkan resolusi spasial, kemudian dilakukan koreksi elevasi menggunakan data *Shuttle Radar Topography Mission* (SRTM) pada variabel suhu. Model dioptimasi menggunakan GridSearchCV dengan validasi silang 5-fold. Hasil evaluasi menunjukkan bahwa model menghasilkan *accuracy* sebesar 0,89, *precision* sebesar 0,89 untuk kedua kelas, *recall* sebesar 0,82 pada kelas 0 serta *F1-score* sebesar 0,85 pada kelas 0 dan 0,91 pada kelas 1. Model ini memberikan kinerja yang lebih baik pada rentang waktu yang lebih panjang, yaitu mulai 09 Desember 2025 hingga 08 Juni 2026. Analisis *feature importance* mengidentifikasi variabel *surface short-wave radiation downwards* (ssrd) sebagai prediktor yang paling berpengaruh dalam proses klasifikasi. Hasil penelitian ini menunjukkan bahwa kombinasi data reanalisis, *statistical downscaling*, dan random forest dapat digunakan untuk mengidentifikasi kondisi ada dan tidak adanya penyinaran matahari pada skala lokal.

Kata kunci: Akurasi klasifikasi, Random forest, *statistical downscaling*, *sunshine duration*, *surface solar radiation*

## ABSTRACT

TIFFANY ANASTASIA JOCELYN. Statistical Downscaling Modeling of Sunshine Duration Data Using Random Forest at Hikmah Farm Pangalengan. Supervised by MUHAMMAD ASYHAR AGMALARO and AZIZ KUSTIYO.

Sunshine duration is an important agroclimatic parameter that supports crop management and agricultural decision-making. This study aims to develop a local sunshine duration occurrence classification model for Hikmah Farm, Pangalengan, using the Random Forest algorithm, with ERA5 atmospheric data as predictor variables and sunshine duration data from Sinaubumi IPB as the binary target variable. In this study, class 1 represents the presence of sunshine duration during the predicted hour, whereas class 0 represents the absence of sunshine duration during the predicted hour. The ERA5 data were first processed using Inverse Distance Weighted (IDW) interpolation to improve spatial resolution, followed by elevation correction using Shuttle Radar Topography Mission (SRTM) data on the temperature variables. The model was optimized using GridSearchCV with 5-fold

cross-validation. The evaluation results showed that the model achieved an accuracy of 0.89, a precision of 0.89 for both classes, a recall of 0.82 for class 0, and F1-scores of 0.85 for class 0 and 0.91 for class 1. The model demonstrated better performance over a longer observation period, spanning from December 9, 2025, to June 8, 2026. Feature importance analysis identified surface short-wave radiation downwards (SSRD) as the most influential predictor in the classification process. The results indicate that the combination of reanalysis data, statistical downscaling, and the Random Forest algorithm can be effectively used to identify the presence and absence of sunshine duration at the local scale.

**Keywords:** Classification accuracy, random forest, statistical downscaling, sunshine duration, surface solar radiation

@ia.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.



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 *STATISTICAL DOWNSCALING* DATA *SUNSHINE DURATION* MENGGUNAKAN RANDOM FOREST DI HIKMAH FARM PANGALENGAN**

**TIFFANY ANASTASIA JOCELYN**

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:  
Prof. Dr. Imas Sukaesih Sitanggang, S.Si, M.Kom



Judul Skripsi : *Pemodelan Statistical Downscaling data Sunshine Duration menggunakan Random forest di Hikmah Farm Pangalengan*

Nama : Tiffany Anastasia Jocelyn

NIM : G6401221005

Disetujui oleh

Pembimbing 1:

Muhammad Asyhar Agmalaro, S.Si., M.Kom

NIP. 198603312012121001

Pembimbing 2:

Dr. Aziz Kustiyo, S.Si, M.Kom

NIP. 197007191998021001

Diketahui oleh

Ketua Program Sarjana Ilmu Komputer:

Dr. Sony Hartono Wijaya, S.Kom., M.Kom

NIP. 198108092008121002

Tanggal Ujian:

1 Juli 2026

Tanggal Lulus:



## PRAKATA

Puji dan syukur penulis panjatkan kepada Tuhan Yesus Kristus atas segala rahmat, karunia, serta pertolongan-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul “Pemodelan *Statistical Downscaling* Data *Sunshine Duration* Menggunakan Random forest di Hikmah Farm Pangalengan.”

Dalam proses penyusunan skripsi ini, penulis menyadari bahwa banyak pihak yang telah memberikan dukungan, bantuan, serta doa. Oleh karena itu, penulis ingin menyampaikan terima kasih yang sebesar-besarnya kepada:

1. Papa Markus Tambayong dan Mama Tjin Yin Chuw, selaku kedua orang tua penulis, atas doa yang tiada henti, kasih sayang, dukungan moral maupun materi, serta kepercayaan dan pengorbanan yang selalu diberikan sehingga penulis dapat menjalani dan menyelesaikan tahap pendidikan ini.
2. Muhammad Asyhar Agmalaro, S.Si., M.Kom. dan Dr. Aziz Kustiyo, S.Si., M.Kom. selaku dosen pembimbing yang telah memberikan arahan, ilmu, kesabaran, serta motivasi kepada penulis selama proses penyusunan penelitian ini.
3. Gabriella Margaretha Jocelyn dan Gerardus Bryan Vernon, selaku adik penulis, atas semangat, keceriaan, dan dukungan yang selalu menjadi penguat bagi penulis di tengah proses penyusunan penelitian ini.
4. Valencius Rianto yang telah menjadi bagian penting dalam perjalanan tahun akhir penulis. Terima kasih telah menjadi rumah untuk melepas keluh kesah, segala usaha yang diberikan mulai dari waktu, dukungan, dan doa dalam proses penyusunan skripsi ini sampai selesai.
5. Noer Hanifah Suganda, Nisrina Indra Putri Zain, Nasywa Nozumi, dan Maysa Fazila Lubis, selaku teman dekat penulis yang selalu memberikan dukungan, semangat, serta menjadi tempat berbagi cerita dan motivasi selama perjalanan akademik ini.
6. Seluruh keluarga, teman-teman, serta pihak lain yang tidak dapat disebutkan satu per satu yang telah memberikan doa, bantuan, dan dukungan sehingga skripsi ini dapat diselesaikan.

Bogor, Juli 2026

*Tiffany Anastasia Jocelyn*



## DAFTAR ISI

|                                                                   |    |
|-------------------------------------------------------------------|----|
| DAFTAR TABEL                                                      | xi |
| DAFTAR GAMBAR                                                     | xi |
| I PENDAHULUAN                                                     | 1  |
| 1.1 Latar Belakang                                                | 1  |
| 1.2 Rumusan Masalah                                               | 2  |
| 1.3 Tujuan                                                        | 2  |
| 1.4 Manfaat                                                       | 3  |
| 1.5 Ruang Lingkup                                                 | 3  |
| II TINJAUAN PUSTAKA                                               | 4  |
| 2.1 Interpolasi IDW (Inverse Distance Weighted)                   | 4  |
| 2.2 <i>Statistical Downscaling</i>                                | 5  |
| 2.3 Random Forest                                                 | 5  |
| 2.4 <i>Hyperparameter Tuning</i>                                  | 6  |
| 2.5 <i>Feature Importance</i>                                     | 6  |
| 2.6 Variabel yang Memengaruhi <i>Sunshine Duration</i>            | 7  |
| 2.7 Perbandingan Hasil Studi Literatur Prediksi Lama Penyinaran   | 8  |
| III METODE                                                        | 12 |
| 3.1 Data Penelitian                                               | 12 |
| 3.2 Tahapan Penelitian                                            | 13 |
| 3.3 Peralatan Penelitian                                          | 19 |
| IV HASIL DAN PEMBAHASAN                                           | 20 |
| 4.1 Eksplorasi Data                                               | 20 |
| 4.2 Praproses Data                                                | 21 |
| 4.3 Interpolasi IDW                                               | 22 |
| 4.4 Koreksi DEM dari SRTM                                         | 25 |
| 4.5 Evaluasi Pemodelan Klasifikasi Random forest                  | 27 |
| 4.5.1 Evaluasi Model Periode 09 Desember 2025 – 10 Februari 2026  | 27 |
| 4.5.2 Evaluasi Model Periode 09 Desember 2025 sampai 08 Juni 2026 | 28 |
| 4.5.3 Pembahasan Hasil Model                                      | 31 |
| 4.6 <i>Statistical Downscaling</i> pada Pemodelan Random forest   | 31 |
| 4.7 <i>Feature Importance</i> pada Random forest                  | 36 |
| 4.8 Perbandingan Aktual dan Prediksi Random forest                | 38 |
| V SIMPULAN DAN SARAN                                              | 42 |
| 5.1 Simpulan                                                      | 42 |
| 5.2 Saran                                                         | 42 |
| DAFTAR PUSTAKA                                                    | 44 |
| RIWAYAT HIDUP                                                     | 45 |

## DAFTAR TABEL

|    |                                                                                                     |    |
|----|-----------------------------------------------------------------------------------------------------|----|
| 1  | Perbandingan studi literatur prediksi lama penyinaran                                               | 8  |
| 2  | Variabel ERA5 pada pemodelan <i>statistical downscaling</i>                                         | 12 |
| 3  | <i>Hyperparameter</i> yang digunakan pada GridSearchCV                                              | 16 |
| 4  | Kinerja random forest pada data uji 09 Desember 2025 sampai 10 Februari 2026                        | 27 |
| 5  | Kinerja random forest sebelum optimasi parameter pada data uji 09 Desember 2025 sampai 08 Juni 2026 | 28 |
| 6  | Kinerja random forest setelah optimasi parameter pada data uji 08 Desember 2025 sampai 08 Juni 2026 | 29 |
| 7  | <i>False Negative</i> pada Data Testing                                                             | 30 |
| 8  | Kinerja random forest menggunakan PCA pada data uji 08 Desember 2025 sampai 08 Juni 2026            | 37 |
| 9  | <i>False Negative</i> pada Data Prediksi                                                            | 38 |
| 10 | Hasil evaluasi rata-rata error harian aktual vs prediksi                                            | 40 |

## DAFTAR GAMBAR

|    |                                                                                                                                                   |    |
|----|---------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 1  | Tahapan penelitian                                                                                                                                | 14 |
| 2  | Ilustrasi tahapan penelitian                                                                                                                      | 14 |
| 3  | Pola rata-rata lama penyinaran berdasarkan jam periode 09 desember 2025 sampai 08 Juni 2026                                                       | 20 |
| 4  | Total lama penyinaran harian periode 09 desember 2025 sampai 08 juni 2026                                                                         | 21 |
| 5  | Visualisasi hasil interpolasi IDW pada masing-masing fitur pada 09 desember 2025 jam 7 pagi (a) t2m, (b) ssrd, (c) d2m, (d) u10, (e) v10, (f) sp  | 24 |
| 6  | Perbandingan hasil interpolasi grid ERA5 dengan hasil interpolasi IDW variabel surface short-wave radiation downwards (ssrd)                      | 25 |
| 7  | Visualisasi statistical downscaling variabel surface short-wave radiation downwards (ssrd) pukul 11 pagi pada tanggal 26 mei 2026                 | 32 |
| 8  | <i>Visualisasi statistical downscaling variabel surface pressure downwards</i> (sp) pukul 11 pagi pada tanggal 26 mei 2026                        | 33 |
| 9  | Visualisasi statistical downscaling variabel 2m temperature (t2m) yang telah dikoreksi pukul 11 pagi pada tanggal 26 mei 2026                     | 34 |
| 10 | <i>Visualisasi statistical downscaling variabel 2m dewpoint temperature</i> (d2m) yang telah dikoreksi pukul 11 pagi pada tanggal 26 mei 2026     | 34 |
| 11 | <i>Visualisasi statistical downscaling variabel total cloud cover (tcc)</i> yang pukul 11 pagi pada tanggal 26 mei 2026                           | 35 |
| 12 | <i>Feature Importance</i> dari model <i>random forest</i>                                                                                         | 36 |
| 13 | Visualisasi perbandingan data aktual dan prediksi random forest (a) Output model random forest (b) Visualisasi total jam lama penyinaran per hari | 39 |