

**PERBANDINGAN *RANDOM FOREST REGRESSION* DAN
SUPPORT VECTOR REGRESSION DALAM PREDIKSI *HOTSPOT*
DI RIAU DENGAN PENDEKATAN *FEATURE ENGINEERING***

BERTHA NITA NAPITUPULU



**PROGRAM STUDI MATEMATIKA
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
BOGOR
2026**

@Hak cipta milik IPB University

IPB University



PERNYATAAN MENGENAI SKRIPSI DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa skripsi dengan judul “Perbandingan *Random Forest Regression* dan *Support Vector Regression* dalam Prediksi *Hotspot* di Riau dengan Pendekatan *Feature Engineering*” 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

Bertha Nita Napitupulu
G5401221055

@Hak cipta milik IPB University

IPB University



ABSTRAK

BERTHA NITA NAPITUPULU. Perbandingan *Random Forest Regression* dan *Support Vector Regression* dalam Prediksi *Hotspot* di Riau dengan Pendekatan *Feature Engineering*. Dibimbing oleh SRI NURDIATI dan ELIS KHATIZAH.

Kebakaran hutan dan lahan di Provinsi Riau merupakan ancaman tahunan yang memerlukan sistem peringatan dini berbasis prediksi *hotspot*. Penelitian ini membandingkan kinerja *Random Forest* (RF) dan *Support Vector Regression* (SVR) dalam memprediksi jumlah *hotspot* menggunakan pendekatan *feature engineering* dengan metrik evaluasi MAE dan RMSE. Peubah yang digunakan meliputi *hotspot*, curah hujan, anomali curah hujan, hari tanpa hujan, ENSO, dan IOD. Didapatkan hasil penerapan *feature engineering* efektif dalam meningkatkan performa *Random Forest* dan *Support Vector Regression* dengan penurunan nilai MAE dan RMSE yang signifikan. Melalui perbandingan kinerja pada kedua model didapatkan kinerja *Random Forest* lebih unggul dibandingkan *Support Vector Regression* yang ditunjukkan oleh nilai MAE dan RMSE yang lebih kecil.

Kata kunci: *feature engineering, hotspot, machine learning, Random Forest, Support Vector Regression*

ABSTRACT

BERTHA NITA NAPITUPULU. Performance Comparison of Random Forest Regression and Support Vector Regression in Hotspot Prediction in Riau Province using a Feature Engineering Approach. Supervised by SRI NURDIATI and ELIS KHATIZAH.

Forest and land fires in Riau Province are an annual threat that requires an early warning system based on hotspot prediction. This study compares the performance of Random Forest (RF) and Support Vector Regression (SVR) in predicting the number of hotspots using a feature engineering approach, evaluated with MAE and RMSE metrics. The variables used include hotspots, rainfall, rainfall anomalies, consecutive dry days, ENSO, and IOD. The results show that the application of feature engineering is effective in improving the performance of both Random Forest and Support Vector Regression, as indicated by a significant reduction in MAE and RMSE values. Based on the performance comparison, Random Forest outperforms Support Vector Regression, as evidenced by lower MAE and RMSE values.

Keywords: *feature engineering, hotspot, machine learning, Random Forest, Support Vector Regression*



@Hak cipta milik IPB University

IPB University

IPB University
Bogor Indonesia



- 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.

**PERBANDINGAN *RANDOM FOREST REGRESSION* DAN
SUPPORT VECTOR REGRESSION DALAM PREDIKSI *HOTSPOT*
DI RIAU DENGAN PENDEKATAN *FEATURE ENGINEERING***

BERTHA NITA NAPITUPULU

Skripsi
sebagai salah satu syarat untuk memperoleh gelar
Sarjana Matematika pada
Program Studi S1 Matematika

**PROGRAM STUDI MATEMATIKA
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
BOGOR
2026**

- 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 University

Tim Penguji pada Ujian Skripsi:
1 Mohamad Khoirun Najib S.Si., M.Mat.

- 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 *Random Forest Regression* dan *Support Vector Regression* dalam Prediksi *Hotspot* di Riau dengan Pendekatan *Feature Engineering*

Nama : Bertha Nita Napitupulu
NIM : G5401221055

Disetujui oleh

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

Pembimbing 2:
Elis Khatizah S.Si., M.Si., Ph.D.

Diketahui oleh

Ketua Program Studi:
Dr. Donny Citra Lesmana, S.Si., M.Fin.Math.
197902272005011001

Tanggal Ujian:
26 Mei 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 September 2025 sampai bulan Mei 2026 ini ialah *machine learning* dengan judul “Perbandingan *Random Forest Regression* dan *Support Vector Regression* dalam Prediksi *Hotspot* di Riau dengan Pendekatan *Feature Engineering*”.

Terima kasih penulis ucapkan kepada:

1. Diri penulis sendiri yang telah berusaha, berproses, dan tidak menyerah dalam menyelesaikan skripsi ini.
2. Ibu Sri Nurdiati dan Ibu Elis Khatizah selaku dosen pembimbing yang telah memberi arahan, bimbingan, serta masukan yang sangat berharga selama proses penyusunan skripsi ini.
3. Orang tua dan keluarga yang senantiasa memberi doa, dukungan, dan semangat kepada penulis.
4. Kak Lilis selaku asisten dosen pembimbing yang telah membantu selama proses penyusunan skripsi ini.
5. Gmatch, Komces, Salu, Dinda Dwi, Dinda Rahma, Fiorenza, Senbuderz, Cathleen, Natasya selaku sahabat penulis yang telah memberi dukungan dan semangat selama proses penyusunan skripsi ini.
6. Teman-teman dan semua pihak yang tidak dapat dapat disebutkan satu per satu yang telah membantu dalam penyelesaian skripsi ini.

Semoga karya ilmiah ini bermanfaat bagi pihak yang membutuhkan dan bagi kemajuan ilmu pengetahuan baik di bidang Matematika, Statistika, maupun Ilmu Komputer.

Bogor, Juli 2026

Bertha Nita Napitupulu

DAFTAR ISI

DAFTAR TABEL	x
DAFTAR GAMBAR	x
DAFTAR LAMPIRAN	xi
I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Tujuan	2
1.4 Manfaat Penelitian	2
II TINJAUAN PUSTAKA	4
2.1 <i>Hotspot</i> sebagai Indikator Kebakaran Hutan	4
2.2 <i>El Nino Southern Oscillation</i> (ENSO)	4
2.3 <i>Indian Ocean Dipole</i> (IOD)	4
2.4 Curah Hujan, Hari Tanpa Hujan dan Anomali Curah Hujan	5
2.5 <i>Random Forest Regression</i>	5
2.6 <i>Support Vector Regression</i> (SVR)	7
2.7 <i>Time Series Cross Validation</i>	9
2.8 <i>Hyperparameter</i>	9
2.9 <i>Random Search</i> dan <i>Grid Search</i>	9
2.10 <i>Autocorrelation Function</i> (ACF)	10
2.11 <i>Feature Engineering</i>	10
2.12 Metrik Evaluasi	11
III METODE PENELITIAN	12
3.1 Sumber dan Jenis Data	12
3.2 Tahapan Penelitian	12
3.2.1 Persiapan dan Pengumpulan Data	12
3.2.2 Ekstraksi dan Pra-pemrosesan Data	12
3.2.3 <i>Feature Engineering</i>	13
3.2.4 Pembagian Data Latih dan Data Uji	14
3.2.5 Pembangunan Model	15
3.2.6 Evaluasi Model	16
IV HASIL DAN PEMBAHASAN	18
4.1 Eksplorasi dan Pra-Pemrosesan Data	18
4.2 Pembangunan Model Tanpa <i>Feature Engineering</i>	20
4.3 <i>Feature Engineering</i>	22
4.3.1 <i>Cyclical Encoding</i>	22
4.3.2 <i>Time-Lag</i> Variabel	23
4.3.3 <i>Feature Interaction</i>	25
4.3.4 <i>Threshold Binary Indicator Variable</i>	25
4.3.5 <i>Rolling Window</i>	26
4.3.6 <i>Feature Selection</i>	27

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.



4.4	Evaluasi Model	27
V	SIMPULAN DAN SARAN	29
5.1	Simpulan	29
5.2	Saran	29
	DAFTAR PUSTAKA	30
	LAMPIRAN	33
	RIWAYAT HIDUP	38

DAFTAR TABEL

1	Fungsi Kernel	9
2	Spesifikasi data yang digunakan	12
3	Visualiasi pembagian <i>cross validation</i>	15
4	Rentang parameter yang digunakan pada pembangunan model <i>random forest</i>	15
5	Rentang parameter yang digunakan pada pembangunan model <i>support vector regression</i>	16
6	Peubah yang digunakan dalam penelitian	18
7	Data jumlah <i>hotspot</i> dan variabel iklim di Riau selama periode Januari 2001 hingga Desember 2024	18
8	Data peubah target setelah ditransformasi	20
9	Data peubah target setelah dinormalisasi	20
10	Parameter terbaik <i>random forest</i> dengan transformasi target	20
11	Parameter terbaik <i>support vector regression</i>	21
12	Hasil evaluasi <i>random forest</i> dan <i>support vector regression</i> tanpa <i>feature engineering</i>	21
13	Transformasi variabel bulan dengan sinus dan osinus	22
14	<i>Lag</i> yang digunakan	25
15	Daftar peubah setelah <i>feature engineering</i>	26
16	Variabel terbaik model	27
17	Kinerja model setelah <i>feature engineering</i>	27

DAFTAR GAMBAR

18	Ilustrasi algoritma <i>random forest</i> (Hana, 2025)	6
19	Ilustrasi ϵ -insentive loss dan <i>slack variable</i> (Smola dan Scholkopf, 2003)	8
20	Grafik jumlah <i>hotspot</i> Riau	18

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.

21	Distribusi data jumlah <i>hotspot</i> Riau	19
22	Distribusi data jumlah <i>hotspot</i> Riau setelah ditransformasi menggunakan \log_{1p}	19
23	Ilustrasi perbandingan hasil prediksi (a) <i>random forest</i> dan (b) <i>support vector regression</i>	21
24	Visualisasi <i>cyclic encoding</i>	23
25	Grafik ACF (a) <i>hotspot</i> (b) hari tanpa hujan (c) curah hujan (d) anomali curah hujan (e) ENSO indeks (f) IOD indeks	24
26	Visualisasi perbandingan hasil prediksi <i>random forest</i> dan <i>support vector regression</i> dengan <i>feature engineering</i>	28

DAFTAR LAMPIRAN

27	Kode program <i>import library</i> dan persiapan model <i>random forest</i>	33
28	Kode pendefinisian variabel baru <i>feature engineering</i>	33
29	Kode Pembagian dataset	34
30	Kode pengisian nilai kosong menggunakan mean data latih	34
31	Kode pembangunan, pelatihan, pengujian model <i>random forest</i>	35
32	Kode program <i>import library</i> dan persiapan model <i>support vector regression</i>	35
33	Kode normalisasi data dan pengisian nilai kosong pada <i>support vector regression</i>	36
34	Kode pembangunan, pelatihan, pengujian model <i>support vector regression</i>	36

