

PERBANDINGAN KINERJA *BEST SUBSET-BAYESIAN* *MODEL AVERAGING* DAN METODE REGULARISASI UNTUK PREDIKSI *HOTSPOT* DI KALIMANTAN

ISNAYNI FEBY HAWARI



**DEPARTEMEN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
INSTITUT PERTANIAN BOGOR
BOGOR
2024**

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 mengumumkannya 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 Kinerja *Best Subset-Bayesian Model Averaging* dan Metode Regularisasi untuk Prediksi *Hotspot* di Kalimantan” 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 2024

Isnayni Feby Hawari
G5401201026

ABSTRAK

ISNAYNI FEBY HAWARI. Perbandingan Kinerja *Best Subset-Bayesian Model Averaging* dan Metode Regularisasi untuk Prediksi *Hotspot* di Kalimantan. Dibimbing oleh SRI NURDIATI dan I WAYAN MANGKU.

Luas wilayah hutan di Kalimantan terus berkurang akibat terjadinya kebakaran hutan dan lahan (karhutla). Penelitian-penelitian sebelumnya menunjukkan bahwa pencegahan karhutla di Kalimantan dapat dilakukan dengan memprediksi jumlah *hotspot* menggunakan *machine learning* berdasarkan indikator iklim. Penelitian ini bertujuan untuk membangun model regresi regularisasi dan *Bayesian Model Averaging* (BMA) berbasis model regresi polinomial berdasarkan hasil *best subset selection*. Model terbaik pada penelitian ini dipilih berdasarkan metrik evaluasi RMSE dan R^2 . Selain itu, pengaruh normalisasi dan standardisasi juga diukur terhadap model yang dibangun. Hasilnya diperoleh kombinasi enam variabel prediktor terbaik dan diperoleh model terbaik yaitu BMA yang merupakan gabungan dari enam model regresi polinomial. Selain itu, penerapan normalisasi dan standardisasi data mempengaruhi nilai parameter seluruh model, tetapi hanya sedikit mempengaruhi hasil pengujian model. Dengan demikian, model BMA dengan data asli lebih dipilih karena nilai koefisiennya lebih mendekati nol dengan performa yang tidak jauh berbeda.

Kata kunci: *best subset selection*, BMA, *hotspot*, regresi, regularisasi

ABSTRACT

ISNAYNI FEBY HAWARI. Performance Comparison of Best Subset-Bayesian Model Averaging and Regularization Method in Predicting the Number of Hotspots in Kalimantan. Supervised by SRI NURDIATI and I WAYAN MANGKU.

The area of forests in Kalimantan continues decreasing due to forest and land fires. Previous studies have shown that the prevention of forest and land fires in Kalimantan can be implemented by predicting the number of hotspots using machine learning based on climate indicators. This study aims to build regularized regression models and Bayesian Model Averaging (BMA) based on polynomial regression models based on best subset selection. The best model in this study is assessed by using the RMSE and R^2 evaluation metrics. In addition, the effect of normalization and standardization is also evaluated on the constructed models. The results shows that a combination of the six best predictor variables is the best and it also shows that the best model is BMA which is a combination of the six polynomial regression models. In addition, the application of data normalization and standardization affects the parameter values of all models, but slightly affects the model testing results. Therefore, the BMA model with original data is more preferable because its coefficients are much closer to zero with similar performance.

Keywords: best subset selection, BMA, hotspot, regression, regularization

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 2024
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 KINERJA *BEST SUBSET-BAYESIAN* *MODEL AVERAGING* DAN METODE REGULARISASI UNTUK PREDIKSI *HOTSPOT* DI KALIMANTAN

ISNAYNI FEBY HAWARI

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

**DEPARTEMEN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
INSTITUT PERTANIAN BOGOR
BOGOR
2024**



@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.

Penguji pada Ujian Skripsi:
Dr. Ir. Retno Budiarti, M.S.

Judul Skripsi : Perbandingan Kinerja *Best Subset-Bayesian Model Averaging* dan Metode Regularisasi untuk Prediksi *Hotspot* di Kalimantan

Nama : Isnayni Feby Hawari

NIM : G5401201026

Disetujui oleh

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



Pembimbing 2:
Prof. Dr. Ir. I Wayan Mangku, M.Sc.



Diketahui oleh

Ketua Departemen Matematika:
Dr. Ir. Endar Hasafah Nugrahani, M.S.
NIP 196312281989032001



Tanggal Ujian: 12 Juni 2024

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 September 2023 sampai bulan April 2024 ini ialah *machine learning*, dengan judul “*Best Subset-Bayesian Model Averaging* dan Model Regularisasi untuk Prediksi Jumlah *Hotspot* di Kalimantan”.

Terima kasih penulis ucapkan kepada:

1. Ibu Prof. Dr. Ir. Sri Nurdiati, M.Sc. selaku dosen pembimbing I dan Bapak Prof. Dr. Ir. I Wayan Mangku, M.Sc. selaku dosen pembimbing II yang telah meluangkan waktu untuk membimbing penulis dalam menyelesaikan penelitian ini hingga akhir,
2. Ibu Dr. Ir. Retno Budiarti, M.S. selaku dosen penguji pada sidang tugas akhir yang telah memberikan kritik dan saran kepada penulis sehingga dapat memperbaiki penelitian ini,
3. Bapak Dadang dan Ibu Neneng Nurindah Sari selaku orang tua penulis, Annisa Aulia Hawari, S.Gz. selaku kakak penulis, dan seluruh keluarga besar yang telah mendoakan serta mendukung penulis hingga mampu menyelesaikan penelitian ini hingga akhir,
4. Mohamad Khoirun Najib selaku asisten dosen pembimbing yang telah membantu membimbing penulis dalam menyelesaikan penelitian ini hingga akhir bersama dengan seluruh dosen pembimbing,
5. Seluruh dosen Departemen Matematika IPB yang telah memberikan ilmu kepada penulis,
6. Seluruh staf Departemen Matematika IPB yang telah membantu penulis dalam mengurus keperluan selama menempuh pendidikan S1 di IPB,
7. Daryl, Elfina, Fanny, Avina, Banissa, Yohana, Sapto, Syukri, Faraj, Izza, dan Farhan selaku teman-teman dekat penulis selama menempuh pendidikan S1 Matematika di IPB yang telah memberikan support kepada penulis sehingga dapat menyelesaikan penelitian ini hingga akhir,
8. Seluruh teman mahasiswa program S1 Departemen Matematika IPB angkatan 57 yang telah memberikan pengalaman dan memori berharga kepada penulis selama menempuh pendidikan S1 di IPB, serta
9. Seluruh pihak lain yang telah membantu penulis menyelesaikan penelitian ini yang tidak dapat disebutkan satu per satu.

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

Bogor, Juli 2024

Isnayni Feby Hawari

DAFTAR ISI

DAFTAR TABEL	x
DAFTAR GAMBAR	x
DAFTAR LAMPIRAN	xi
I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Tujuan	3
II TINJAUAN PUSTAKA	4
2.1 <i>Hotspot</i>	4
2.2 Indikator Iklim	4
2.3 Transformasi Data	5
2.4 Regresi Linear Berganda	6
2.5 <i>Best Subset Selection</i>	7
2.6 Regresi Regularisasi	7
2.7 <i>Bayesian Model Averaging (BMA)</i>	9
2.8 <i>Bayesian Information Criterion (BIC)</i>	10
2.9 Metrik Evaluasi	10
III METODE	12
3.1 Sumber dan Jenis Data	12
3.2 Tahapan Penelitian	12
IV HASIL DAN PEMBAHASAN	16
4.1 Praproses Data	16
4.2 Pelatihan Model	20
4.3 Pengujian Model	23
4.4 Analisis Pengaruh Normalisasi dan Standardisasi Data	24
V SIMPULAN DAN SARAN	29
5.1 Simpulan	29
5.2 Saran	29
DAFTAR PUSTAKA	30
LAMPIRAN	33
RIWAYAT HIDUP	39

DAFTAR TABEL

1	Variabel yang digunakan dalam penelitian	16
2	Variabel prediktor tambahan hasil penerapan <i>polynomial features</i>	18
3	Nilai <i>hyperparameter</i> yang digunakan pada setiap model regularisasi	20
4	Nilai <i>intercept</i> dan koefisien regresi untuk setiap variabel prediktor pada ketiga model regularisasi	21
5	Nilai parameter model BMA pada setiap model regresi polinomial terpilih	22
6	Nilai RMSE dan <i>R-Squared</i> pada data latih dan uji untuk setiap model	24
7	Perbandingan nilai data asli, normalisasi, dan standardisasi pada dua contoh variabel prediktor untuk periode Januari 2001 hingga Mei 2001	25
8	Nilai RMSE dan <i>R-squared</i> setiap model pada data normalisasi dan standardisasi	28

DAFTAR GAMBAR

1	Alur tahapan penelitian	13
2	<i>Heatmap</i> nilai korelasi antar variabel	17
3	Hasil seleksi variabel menggunakan <i>best subset selection</i>	19
4	Perbandingan performa model regresi <i>ridge</i> , LASSO, <i>elastic-net</i> , dan model akhir BMA dengan nilai observasi	23
5	Perbandingan koefisien model regresi regularisasi pada data asli, normalisasi, dan standardisasi	26
6	Perbandingan nilai koefisien model akhir BMA berbasis model regresi polinomial pada data asli, normalisasi, dan standardisasi	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 mengumumkannya dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

DAFTAR LAMPIRAN

1. <i>Library</i> bahasa pemrograman R yang digunakan	34
2. Algoritma penerapan <i>polynomial features</i> dengan derajat polinomial maksimum 3	34
3. Algoritma seleksi variabel menggunakan <i>best subset selection</i>	34
4. Algoritma pelatihan salah satu model regresi regularisasi yaitu regresi <i>ridge</i>	35
5. Algoritma pelatihan model BMA berbasis model regresi polinomial	36
6. Algoritma pengujian salah satu model yaitu model regresi <i>ridge</i>	36
7. Hasil uji asumsi model terbaik pada data asli	37
8. Hasil uji asumsi multikolinearitas menggunakan nilai <i>Variance Inflation Factor</i> (VIF) model terbaik pada data asli	37
9. Nilai <i>intercept</i> dan koefisien model regresi polinomial terpilih untuk model BMA pada data normalisasi	37
10. Nilai <i>intercept</i> dan koefisien model regresi polinomial terpilih untuk model BMA pada data standardisasi	38