

PERBANDINGAN PERFORMA METODE *RANDOM FOREST* DAN MARS PADA PENGKLASIFIKASIAN STATUS RAWAN PANGAN DI PROVINSI BANTEN TAHUN 2023

ROSY ROSITA



**PROGRAM STUDI STATISTIKA DAN SAINS DATA
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
BOGOR
2025**



@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 “Perbandingan Performa Metode *Random Forest* dan MARS pada Pengklasifikasian Status Rawan Pangan di Provinsi Banten Tahun 2023” 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, Januari 2025

Rosy Rosita
G1401201016

- 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

ROSY ROSITA. Perbandingan Performa Metode *Random Forest* dan MARS pada Pengklasifikasian Status Rawan Pangan di Provinsi Banten Tahun 2023. Dibimbing oleh KUSMAN SADIK dan ITASIA DINA SULVIANTI.

Kerawanan pangan merupakan kondisi ketidakmampuan wilayah sampai dengan perseorangan yang tercermin dari tidak tersedianya pangan yang cukup. Salah satu upaya yang dapat dilakukan adalah dengan mengoptimalkan tindakan pencegahan dan intervensi berdasarkan status rawan pangan. Data BPS tahun 2023 menunjukkan angka prevalensi rawan pangan di Provinsi Banten mengalami peningkatan dari 4,86 persen pada tahun 2021 menjadi 5,76 persen pada tahun 2023. Penelitian ini bertujuan untuk memodelkan kejadian rawan pangan rumah tangga dengan menerapkan metode *Multivariate Adaptive Regression Spline* (MARS) sebagai pengklasifikasi tunggal dan *random forest* sebagai metode klasifikasi *ensemble* pada data rawan pangan di Provinsi Banten. Data yang digunakan merupakan data Survey Sosial Ekonomi Nasional tahun 2023, data berjumlah 7160 data dengan unit penelitiannya adalah rumah tangga. Hasil penelitian menunjukkan bahwa model *Synthetic Minority Oversampling Technique* (SMOTE) *random forest* dan SMOTE MARS memiliki perbedaan performa yang tidak berbeda nyata. Kedua model menghasilkan dua peubah terpenting yang sama yaitu tingkat pendidikan terakhir yang ditempuh oleh kepala rumah tangga dan besar pangsa pengeluaran pangan.

Kata kunci: kerawanan pangan, *multivariate adaptive regression spline*, pemodelan klasifikasi, *random forest*



ABSTRACT

ROSY ROSITA. Performance Comparison of Random Forest and MARS Method in Classifying Food Insecurity Status in Banten Province 2023. Supervised by KUSMAN SADIK and ITASIA DINA SULVIANTI.

Food insecurity is a condition of regional to individual inability reflected in the unavailability of sufficient food. One of the ways that can be done is to optimize prevention and intervention measures based on food insecurity status. BPS data for 2023 shows that the prevalence rate of food insecurity in Banten Province has increased from 4.86 percent in 2021 to 5.76 percent in 2023. This study aims to model the incidence of household food insecurity by applying the Multivariate Adaptive Regression Spline (MARS) method as a single classifier and random forest as an ensemble classification method to food insecurity data in Banten Province. The data used is Survey Sosial Ekonomi Nasional data in 2023, 7160 data with the research unit is households. The results showed that the Synthetic Minority Oversampling Technique (SMOTE) random forest and SMOTE MARS models have statistically insignificant differences in performance. Both models produced the same two most important variables, namely the level of education attained by the household head and the share of food expenditure.

Keywords: classification model, food insecurity, multivariate adaptive regression spline, random forest



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 2025
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 PERFORMA METODE *RANDOM FOREST* DAN MARS PADA PENGKLASIFIKASIAN STATUS RAWAN PANGAN DI PROVINSI BANTEN TAHUN 2023

ROSY ROSITA

Skripsi
sebagai salah satu syarat untuk memperoleh gelar
Sarjana pada
Program Studi Statistika dan Sains Data

**PROGRAM STUDI STATISTIKA DAN SAINS DATA
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
BOGOR
2025**



@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.



@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 : Perbandingan Performa Metode *Random Forest* dan MARS pada Pengklasifikasian Status Rawan Pangan di Provinsi Banten Tahun 2023

Nama : Rosy Rosita
NIM : G1401201016

Disetujui oleh

Pembimbing 1:
Dr. Kusman Sadik S.Si., M.Si.



Pembimbing 2:
Dra. Itasia Dina Sulvianti M.Si.



Diketahui oleh

Ketua Program Studi Statistika dan Sains Data:
Dr. Bagus Sartono S.Si., M.Si.
NIP 197804112005011002



Tanggal Ujian:
17 Desember 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 Maret 2024 sampai bulan Oktober 2024 ini ialah klasifikasi, dengan judul “Perbandingan Performa Metode *Random Forest* dan MARS pada Pengklasifikasian Status Rawan Pangan di Provinsi Banten Tahun 2023”.

Karya ilmiah ini dapat diselesaikan tidak terlepas dari dukungan berbagai pihak. Oleh karena itu, penulis ingin mengucapkan rasa terima kasih diantara kepada:

1. Bapak Dr. Kusman Sadik, S.Si., M.Si. dan Ibu Dra. Itasia Dina Sulvianti, M.Si., selaku dosen pembimbing yang telah senantiasa membimbing, memberi saran dan masukan dalam penulisan karya ilmiah ini.
2. Bapak/Ibu selaku dosen penguji yang telah memberikan kritik dan saran dalam penulisan karya ilmiah ini.
3. Badan Pusat Statistik yang telah mengizinkan penggunaan data pada penulisan karya ilmiah ini.
4. Seluruh dosen pengajar Departemen Statistika IPB yang telah memberikan penulis banyak ilmu dan pembelajaran bermanfaat selama perkuliahan.
5. Orang tua serta keluarga penulis yang senantiasa memberikan doa, dukungan, nasihat, kasih sayang, dan kenyamanan bagi penulis.
6. Seluruh Staff Tata Usaha Departemen Statistika IPB yang telah banyak membantu penulis dalam proses administrasi perkuliahan.
7. Teman-teman Statistika Angkatan 57 dan teman-teman penulis lainnya yang telah memberikan banyak bantuan, saran, motivasi, serta semangat selama masa perkuliahan hingga penyusunan karya ilmiah ini.
8. Seluruh pihak terkait yang tidak dapat penulis sebutkan satu persatu atas segala doa, kebaikan, dukungan, serta motivasi sehingga penulis dapat menyelesaikan karya ilmiah ini.

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

Bogor, Januari 2025

Rosy Rosita

DAFTAR ISI

DAFTAR TABEL	x
DAFTAR GAMBAR	x
DAFTAR LAMPIRAN	xi
I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Tujuan	2
II TINJAUAN PUSTAKA	3
2.1 Kerawanan Pangan	3
2.2 Data Tidak Seimbang	3
2.3 <i>Synthetic Minority Oversampling Technique (SMOTE)</i>	4
2.4 <i>Metode Ensemble</i>	4
2.5 <i>Random Forest</i>	5
2.6 <i>Multivariate Adaptive Regression Spline (MARS)</i>	7
2.7 <i>K-Fold Cross Validation</i>	10
2.8 <i>Grid Search</i>	10
2.9 Evaluasi Kebaikan Model	11
2.10 Uji <i>T</i> Berpasangan	13
III METODE	15
3.1 Data	15
3.2 Prosedur Analisis Data	16
IV HASIL DAN PEMBAHASAN	19
4.1 Eksplorasi Karakteristik Rumah Tangga di Provinsi Banten	19
4.2 Pembagian Data Latih dan Data Uji	21
4.3 Penanganan Data Tidak Seimbang	22
4.4 Pemodelan Klasifikasi <i>Random Forest</i>	23
4.5 Pemodelan Klasifikasi <i>Multivariate Adaptive Regression Splines (MARS)</i>	25
4.6 Perbandingan Hasil Klasifikasi MARS dan <i>Random Forest</i>	29
V SIMPULAN DAN SARAN	33
5.1 Simpulan	33
5.2 Saran	33
DAFTAR PUSTAKA	34
LAMPIRAN	39
RIWAYAT HIDUP	49



DAFTAR TABEL

1	Matriks konfusi	11
2	Kategori keakuratan hasil klasifikasi berdasarkan nilai AUC	13
3	Peubah yang digunakan dalam penelitian	15
4	Proporsi awal data latih dan data uji	21
5	Persentase awal perbandingan data jumlah rumah tangga rawan pangan	22
6	Persentase perbandingan data jumlah rumah tangga rawan pangan setelah dilakukan penanganan SMOTE	22
7	Perbandingan performa model <i>random forest</i> sebelum dan setelah SMOTE dalam setiap <i>fold</i> pada data latih	23
8	Uji <i>t</i> rata-rata performa model <i>random forest</i> sebelum dan setelah SMOTE	24
9	Perbandingan performa model MARS sebelum dan setelah SMOTE dalam setiap <i>fold</i> pada data latih	25
10	Uji <i>t</i> rata-rata performa model MARS sebelum dan setelah SMOTE	26
11	Model SMOTE MARS	26
12	Perbandingan performa SMOTE MARS dan SMOTE <i>random forest</i> pada data latih	30
13	Uji <i>t</i> rata-rata performa model SMOTE MARS dan SMOTE <i>random forest</i>	30
14	Perbandingan performa pada data uji	31

DAFTAR GAMBAR

1	Ilustrasi metode <i>ensemble</i>	5
2	Ilustrasi <i>random forest</i>	6
3	Ilustrasi metode MARS	7
4	Ilustrasi <i>k-fold cross validation</i>	10
5	Ilustrasi <i>grid search</i>	11
6	Visualisasi dari kurva ROC	12
7	Diagram alir penelitian	18
8	Proporsi rumah tangga rawan pangan di Provinsi Banten	19
9	Peta persentase persebaran rumah tangga rawan pangan berdasarkan kabupaten/kota	20
10	Plot persentase rumah tangga rawan pangan berdasarkan peubah penjas	20
11	Grafik sebelum dan sesudah dilakukan penanganan ketidakseimbangan SMOTE	22
12	Tingkat kepentingan peubah berdasarkan model SMOTE <i>random forest</i>	24
13	Tingkat kepentingan peubah berdasarkan model SMOTE MARS	29

14	Perbandingan performa pada data latih dan data uji untuk model (a) SMOTE MARS dan (b) SMOTE <i>Random Forest</i>	31
----	--	----

DAFTAR LAMPIRAN

1	Pertanyaan FIES pada SUSENAS	40
2	Persentase status rawan pangan berdasarkan kategori peubah penjelas	41
3	Model MARS pada Data Tanpa Penanganan Ketidakseimbangan SMOTE	42
4	Tingkat kepentingan peubah berdasarkan model MARS	42
5	<i>Mean Decrease Gini</i> pada model klasifikasi <i>random forest</i>	43
6	Hasil <i>tuning</i> MARS	43
7	Hasil <i>tuning</i> SMOTE MARS	43
8	Grafik hasil <i>tuning</i> SMOTE MARS	44
9	Hasil <i>tuning random forest</i>	45
10	Hasil <i>tuning</i> SMOTE <i>random forest</i>	45
11	Grafik hasil <i>tuning</i> SMOTE <i>random forest</i>	46
12	Uraian performa <i>F1-score</i> kedua model	46
13	Nilai GCV pada setiap tahapan eliminasi BF	47
14	Nilai kumulatif penurunan GCV pada setiap tahapan eliminasi BF untuk masing-masing peubah penjelas	48