

EVALUASI METODE *HYBRID FAY-HERRIOT* RANDOM FOREST DALAM PREDIKSI PENGELUARAN PER KAPITA DI KABUPATEN BANDUNG

NABILLAH RAHMATIAH TANGKE



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

**BOGOR
2026**





PERNYATAAN MENGENAI TESIS DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa tesis dengan judul “Evaluasi Metode *Hybrid Fay-Herriot Random Forest* dalam Prediksi Pengeluaran Per Kapita di Kabupaten Bandung” 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 tesis ini.

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

Bogor, Agustus 2026

Nabillah Rahmatiah Tangke
M0501241070



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.



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 mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

IPB University

RINGKASAN

NABILLAH RAHMATIAH TANGKE. Evaluasi Metode *Hybrid Fay-Herriot Random Forest* dalam Prediksi Pengeluaran Per Kapita di Kabupaten Bandung. Dibimbing oleh ASEP SAEFUDDIN dan ANANG KURNIA.

Penduga langsung dari survei pada area kecil sering kali tidak reliabel akibat keterbatasan ukuran contoh. Pendekatan *Small Area Estimation* (SAE) berbasis model Fay-Herriot mampu menstabilkan penduga melalui mekanisme borrowing strength, namun metode ini bertumpu pada asumsi linieritas antara peubah respon dan peubah penyerta. Pada data sosial ekonomi yang sarat pola non-linier dan interaksi kompleks, pemaksaan asumsi linieritas berisiko menyebabkan salah-spesifikasi model yang memperburuk ketepatan prediksi.

Penelitian ini bertujuan untuk mengembangkan dan menerapkan model hibrida Fay-Herriot Random Forest (FH-RF) yang menggantikan komponen linier $X'\beta$ pada model Fay-Herriot dengan fungsi non-linier $f(X)$ dari algoritma *random forest* guna meningkatkan akurasi pendugaan rata-rata pengeluaran per kapita bahan makanan, minuman, dan rokok pada tingkat kecamatan di Kabupaten Bandung. Data empiris bersumber dari Survei Indeks Pembangunan Manusia (IPM) Kabupaten Bandung dengan 31 kecamatan sebagai unit area kecil dengan ukuran contoh 15-20, serta data Potensi Desa (PODES) sebagai sumber 10 peubah penyerta yang mencakup indikator infrastruktur ekonomi, aksesibilitas fasilitas publik, dan kondisi kesejahteraan. Penduga parameter ragam efek acak dilakukan melalui Algoritma Expectation-Maximization berdasarkan pendekatan Luisávilá-Valdez *et al.* (2020), sedangkan penduga MSE pada model hibrida menggunakan Parametric Bootstrap. Evaluasi performa dilakukan secara komprehensif melalui sembilan skenario pemodelan yang mengkombinasikan tiga aspek yaitu jenis model dimana model yang digunakan adalah Fay-Herriot klasik vs Fay-Herriot Random Forest, transformasi respons (skala asli vs log-normal), dan sumber ragam penarikan contoh (ragam langsung vs ragam terkoreksi Generalized Variance Function).

Hasil analisis menunjukkan bahwa model hibrida FH-RF mengungguli model Fay-Herriot klasik maupun penduga langsung pada seluruh metrik evaluasi. Skenario terbaik, yaitu Hybrid FH-RF Log-Normal dengan ragam GVF, menghasilkan rata-rata Relative Root Mean Square Error (RRMSE) sebesar 7,14%, jauh lebih rendah dibandingkan model Fay-Herriot (9,05%) maupun penduga langsung (12,75%). Penurunan RRMSE dari model Fay-Herriot mengkonfirmasi bahwa hubungan antara peubah penyerta PODES dan pengeluaran per kapita di Kabupaten Bandung bersifat non-linier, sehingga pendekatan Random Forest yang mampu menangkap pola tersebut memberikan perbaikan substansial. Sebanyak 28 dari 31 kecamatan (90,3%) mencapai RRMSE di bawah 10% pada skenario terbaik, menunjukkan presisi penduga yang memadai untuk seluruh wilayah. Temuan ini memberikan peta pengeluaran yang lebih akurat pada tingkat kecamatan sebagai landasan strategis bagi Pemerintah Kabupaten Bandung dalam merumuskan kebijakan pengentasan kemiskinan yang spesifik dan tepat sasaran sejalan dengan pencapaian SDG 1 Tanpa Kemiskinan.

Kata kunci: EBLUP, *Fay-Herriot*, Pengeluaran Per Kapita, *Random Forest*, Penduga Area Kecil



- 



SUMMARY

NABILLAH RAHMATIAH TANGKE. Evaluation of Hybrid Fay-Herriot Random Forest Methods for Per Capita Expenditure Estimation in Bandung Regency. Supervised by ASEP SAEFUDDIN, and ANANG KURNIA.

Direct estimation from surveys in small areas is often unreliable due to limited sample sizes. The Small Area Estimation (SAE) approach based on the Fay-Herriot model can stabilize estimates through borrowing strength mechanisms; however, this method relies on the assumption of linearity between the response variable and auxiliary variables. In socioeconomic data characterized by nonlinear patterns and complex interactions, imposing the linearity assumption risks model misspecification, which deteriorates prediction accuracy.

This study aims to develop and implement a hybrid Fay-Herriot Random Forest (FH-RF) model that replaces the linear component $X'\beta$ in the Fay-Herriot model with a nonlinear function $f(X)$ estimated by the random forest algorithm to improve the accuracy of estimating average per capita expenditure on food, beverages, and tobacco at the sub-district level in Bandung Regency. Empirical data were sourced from the Human Development Index (HDI) Survey of Bandung Regency with 31 sub-districts as small area units with sample size of 15-20 households per sub-district, and Village Potential Statistics (PODES) data as the source of 10 auxiliary variables encompassing economic infrastructure indicators, public facility accessibility, and welfare conditions. Random effect variance parameters were estimated through the Expectation-Maximization algorithm based on the approach of Luisávilá-Valdez et al. (2020), while MSE estimation for the hybrid model employed parametric bootstrap. Performance evaluation was conducted comprehensively through nine modeling scenarios combining three aspects: model type (classic Fay-Herriot EBLUP-REML vs. hybrid Fay-Herriot Random Forest), response transformation (original scale vs. log-normal), and sampling variance source (direct variance vs. Generalized Variance Function corrected variance).

The results demonstrate that the hybrid FH-RF model consistently outperforms both the Fay-Herriot model and direct estimation across all evaluation metrics. The best-performing scenario, Hybrid FH-RF Log-Normal with GVF variance, yielded a mean Relative Root Mean Square Error (RRMSE) of 7.14%, substantially lower than the best classic Fay-Herriot model (9.05%) and direct estimation (12.75%). The RRMSE reduction from the Fay-Herriot model confirms that the relationship between PODES auxiliary variables and per capita expenditure in Bandung Regency is inherently nonlinear, and the Random Forest approach capable of capturing such patterns, provides substantial improvement. A total of 28 out of 31 sub-districts (90.3%) achieved RRMSE below 10% under the best scenario, indicating adequate estimation precision across the entire study area. These findings provide a more accurate expenditure map at the sub-district level, serving as a strategic foundation for the Bandung Regency Government in formulating policies aligned with the achievement of SDG 1 No Poverty.

Keywords: Small Area Estimation, Fay-Herriot, Random Forest, EBLUP, Per Capita Expenditure, Human Development Index



- 

EVALUASI METODE *HYBRID FAY-HERRIOT RANDOM FOREST* DALAM PREDIKSI PENGELUARAN PER KAPITA DI KABUPATEN BANDUNG

NABILLAH RAHMATIAH TANGKE

Tesis
sebagai salah satu syarat untuk memperoleh gelar
Magister pada
Program Studi Statistika dan Sains Data

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



@Hak cipta milik IPB University

IPB University

Tim Penguji pada Ujian Tesis:
Dr.Ir. Indahwati M.Si

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 Tesis : Evaluasi Metode *Hybrid Fay-Herriot Random Forest* dalam
Prediksi Pengeluaran Per Kapita di Kabupaten Bandung
Nama : Nabillah Rahmatiah Tangke
NIM : M0501241070

Disetujui oleh

Pembimbing 1:
Prof. Dr. Ir. Asep Saefuddin, M.Sc.

Pembimbing 2:
Prof. Dr. Anang Kurnia, S.Si., M.Si.



Diketahui oleh

Ketua Program Studi:
Dr. Agus Mohamad Soleh, S.Si, MT.
NIP 197503151999031004

Dekan Sekolah Sains Data, Matematika, dan Informatika:
Prof. Dr. Ir. Agus Bueno, M.Si., M.Kom.
NIP 19660702 199302 1 001



Tanggal Ujian: 07 Agustus 2026

Tanggal Lulus:



- 



PRAKATA

Puji syukur penulis panjatkan ke hadirat Allah SWT atas segala limpahan rahmat dan karunia-Nya, sehingga tesis yang berjudul “Evaluasi Metode *Hybrid Fay-Herriot Random Forest* dalam Prediksi Pengeluaran Per Kapita di Kabupaten Bandung” ini dapat diselesaikan dengan baik. Penyusunan tesis ini merupakan salah satu tahapan penting dalam pemenuhan syarat kelulusan pada program magister.

Selama proses penyusunan, penulis banyak menerima bantuan, bimbingan, serta dukungan moril maupun materiil dari berbagai pihak. Oleh karena itu, penulis ingin menyampaikan apresiasi dan terima kasih yang tulus kepada:

1. Bapak Prof. Dr. Ir. Asep Saefuddin, M.Sc. dan Bapak Prof. Dr. Anang Kurnia, S.Si., M.Si. selaku dosen pembimbing yang telah meluangkan waktu untuk memberikan arahan, ilmu, serta saran yang sangat berharga dalam penyempurnaan penelitian ini.
2. Seluruh dosen dan tenaga kependidikan Departemen Statistika IPB University atas bekal ilmu dan fasilitas yang sangat mendukung perjalanan akademik penulis.
3. Orang tua, keluarga besar, yang senantiasa memberikan doa, cinta, dan dukungan tanpa henti yang menjadi kekuatan bagi penulis.
4. Teman-teman Program Studi Statistika dan Sains Data Angkatan 2024, terima kasih atas semangat, diskusi, dan kebersamaan yang luar biasa selama ini.

Penulis menyadari bahwa tesis ini masih jauh dari kesempurnaan. Oleh sebab itu, kritik dan saran yang membangun sangat penulis harapkan. Akhir kata, semoga tesis ini dapat memberikan kontribusi nyata bagi perkembangan ilmu pengetahuan serta menjadi referensi yang bermanfaat dalam perumusan kebijakan berbasis data di Indonesia.

Bogor, Agustus 2026

Nabillah Rahmatiah Tangke

DAFTAR ISI

DAFTAR ISI	vii
DAFTAR TABEL	viii
DAFTAR GAMBAR	viii
DAFTAR LAMPIRAN	viii
I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Tujuan	4
1.4 Manfaat	4
II TINJAUAN PUSTAKA	6
2.1 <i>Small Area Estimation</i>	6
2.1.1 Level Area	6
2.1.2 Level Unit	7
2.2 EBLUP	8
2.3 <i>Restricted Maximum Likelihood (REML)</i>	9
2.4 <i>Machine Learning</i>	10
2.5 <i>Fay-Herriot Random Forest (FH-RF)</i>	11
2.6 Model Transformasi Lognormal	12
2.7 <i>Generalized Variance Function</i>	13
2.8 <i>Sustainable Development Goals (SDGs)</i>	14
III METODE	15
3.1 Pendekatan Simulasi	15
3.2 Pendekatan Analisis Data Empiris	17
IV HASIL DAN PEMBAHASAN	20
4.1 Kajian Simulasi	20
4.1.1 Eksplorasi Data Simulasi	20
4.1.2 Pemodelan	25
4.2 Kajian Empiris	28
4.2.1 Eksplorasi Data Empiris	28
4.2.2 Pemodelan	34
V SIMPULAN DAN SARAN	39
5.1 Simpulan	39
5.2 Saran	39
DAFTAR PUSTAKA	40
LAMPIRAN	43
RIWAYAT HIDUP	45

DAFTAR TABEL

1	Skenario pembangkitan data simulasi	15
2	Peubah respon dan prediktor data empiris	17
3	Karakteristik peubah pada tiap-tiap skenario	20
4	Hasil uji shapiro-wilk peubah pada masing-masing skenario	21
5	Nilai VIF masing-masing peubah	25
6	MSE masing-masing model per skenario	25
7	RRMSE masing-masing model per skenario	26
8	Tabel Uji ANOVA 3 Faktor pada RRMSE	28
9	Karakteristik peubah respon dan prediktor	29
10	Diagnostik linieritas dan struktur galat model	32
11	Nilai VIF peubah prediktor	34
12	Hasil evaluasi kajian empiris	35

DAFTAR GAMBAR

1	Tahapan kajian data simulasi.	16
2	Tahapan kajian data empiris.	18
3	Histogram distribusi peubah respon data simulasi.	22
4	Boxplot sebaran peubah prediktor data simulasi.	23
5	<i>Lineplot</i> Residual vs <i>Fitted Values</i> .	24
6	<i>Barplot</i> perbandingan nilai RRMSE masing-masing metode pada tiap-tiap skenario.	27
7	<i>Histogram</i> peubah respon pengeluaran per Kapita makanan dan non makanan, (a) Sebaran peubah respon sebelum dilakukan transformasi log normal, (b) Sebaran peubah respon setelah dilakukan transformasi log normal.	30
8	<i>Histogram</i> peubah prediktor data empiris.	31
9	<i>Heatmap</i> korelasi data empiris.	33
10	<i>Lineplot</i> perbandingan nilai penduga per Kecamatan pada tiap-tiap skenario model.	36
11	<i>Lineplot</i> perbandingan nilai RRMSE per Kecamatan pada tiap-tiap skenario model.	37

DAFTAR LAMPIRAN

1	Lampiran 1 Uji Tukey Interaksi Metode dan Tipe	43
2	Lampiran 2 prediksi per Area masing-masing model	43
3	Lampiran 3 RRMSE per Area masing-masing model	44