

EVALUASI KINERJA MODEL GEOGRAPHICAL-XGBOOST PADA DATA DENGAN TINGKAT HETEROGENITAS SPASIAL DAN POLA LINEARITAS BERBEDA

ZAMRAH MUTMAINAH



**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 Kinerja Model *Geographical-XGBoost* pada Data dengan Tingkat Heterogenitas Spasial dan Pola Linearitas Berbeda” 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, Juli 2026

Zamrah Mutmainah
M0501241066

- 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.

© 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.

RINGKASAN

ZAMRAH MUTMAINAH. Evaluasi Kinerja Model *Geographical-XGBoost* pada Data dengan Tingkat Heterogenitas Spasial dan Pola Linearitas Berbeda Dibimbing oleh BUDI SUSETYO dan SEPTIAN RAHARDIANTORO.

Analisis data spasial sering menghadapi tantangan akibat adanya heterogenitas spasial dan hubungan nonlinear antarpeubah. Metode konvensional seperti *Ordinary Least Squares* (OLS) dan *Geographically Weighted Regression* (GWR) memiliki keterbatasan dalam menangkap kedua karakteristik tersebut secara bersamaan. OLS mengasumsikan hubungan yang bersifat global dan homogen, sedangkan GWR hanya mampu mengakomodasi heterogenitas spasial dengan asumsi hubungan linear. Di sisi lain, metode *Extreme Gradient Boosting* (XGBoost) memiliki kemampuan yang baik dalam memodelkan hubungan nonlinear, namun belum mempertimbangkan variasi spasial. Pengembangan metode *Geographically Weighted-XGBoost* (GW-XGBoost) telah menggabungkan aspek spasial dan nonlinear, tetapi masih menggunakan pendekatan hibrida yang berpotensi menyebabkan *overfitting*. Oleh karena itu, penelitian ini mengevaluasi kinerja metode *Geographical-XGBoost* (G-XGBoost) yang mengintegrasikan pembobotan geografis secara langsung ke dalam algoritma XGBoost.

Penelitian ini bertujuan mengevaluasi performa G-XGBoost dibandingkan dengan OLS, GWR, XGBoost, dan GW-XGBoost pada berbagai tingkat heterogenitas spasial dan pola hubungan linear maupun nonlinear. Evaluasi dilakukan melalui kajian simulasi menggunakan empat skenario yang mengombinasikan hubungan linear dan nonlinear dengan heterogenitas spasial rendah maupun tinggi. Selanjutnya, seluruh metode diterapkan pada data tingkat kemiskinan kabupaten/kota di Pulau Jawa tahun 2024 untuk menentukan model terbaik. Performa model dievaluasi menggunakan koefisien determinasi (R^2), *Root Mean Squared Error* (RMSE), dan *Mean Absolute Error* (MAE), sedangkan stabilitas model dinilai berdasarkan selisih performa antara data latih dan data uji.

Hasil kajian simulasi menunjukkan bahwa G-XGBoost memiliki performa yang paling stabil dan konsisten pada seluruh skenario yang diuji. Pada skenario linear, OLS dan GWR memberikan hasil yang kompetitif, namun performanya menurun pada data nonlinear. Sebaliknya, G-XGBoost mampu mempertahankan akurasi prediksi yang tinggi, baik pada hubungan linear maupun nonlinear, serta pada tingkat heterogenitas spasial yang berbeda. Selain itu, G-XGBoost menunjukkan gap antara data latih dan data uji yang paling kecil dibandingkan metode lain, sehingga memiliki kemampuan generalisasi yang lebih baik dan risiko *overfitting* yang lebih rendah dibandingkan GW-XGBoost.

Pada kajian empiris, G-XGBoost kembali menunjukkan performa terbaik dengan nilai R^2 sebesar 0,8115 dan RMSE sebesar 1,6241, yang merupakan hasil terbaik dibandingkan model pembanding. Analisis *Local Variable Importance* (LVI) menunjukkan bahwa tingkat kepentingan relatif peubah penjelas dalam proses prediksi tingkat kemiskinan berbeda antar kabupaten/kota di Pulau Jawa. Rata-rata lama sekolah merupakan salah satu peubah yang paling sering memiliki nilai kepentingan peubah (*variable importance*) tinggi, sedangkan persentase penduduk yang bekerja di sektor pertanian, persentase rumah tangga dengan alas lantai semipermanen, upah minimum, kepemilikan jaminan kesehatan, dan



persentase penduduk berpendidikan minimal SMA teridentifikasi sebagai peubah dengan *variable importance* tinggi pada wilayah tertentu. Hasil penelitian ini menunjukkan bahwa G-XGBoost mampu mengakomodasi hubungan nonlinear dan heterogenitas spasial secara bersamaan, yang dapat menjadi referensi awal dalam mengidentifikasi peubah-peubah yang memiliki tingkat kepentingan relatif tinggi untuk dikaji lebih lanjut sesuai karakteristik masing-masing wilayah.

Kata kunci: *Geographical-XGBoost*, *Geographically Weighted Regression* (GWR), heterogenitas spasial, kemiskinan, XGBoost.

1. Hak Cipta Dilindungi Undang-undang
 - a. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - b. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - c. 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.

SUMMARY

ZAMRAH MUTMAINAH. Evaluation of Geographical-XGBoost Model Performance on Data with Different Levels of Spatial Heterogeneity and Linearity Patterns. Supervised by BUDI SUSETYO and SEPTIAN RAHARDIANTORO

Spatial data analysis often faces challenges arising from spatial heterogeneity and nonlinear relationships among variables. Conventional methods such as Ordinary Least Squares (OLS) and Geographically Weighted Regression (GWR) are limited in their ability to capture these two characteristics simultaneously. OLS assumes a global, homogeneous relationship across all observations, whereas GWR accommodates spatial heterogeneity while still relying on the assumption of linear relationships. On the other hand, Extreme Gradient Boosting (XGBoost) effectively models nonlinear relationships but does not explicitly account for spatial variation. The development of Geographically Weighted XGBoost (GW-XGBoost) combines spatial and nonlinear modeling capabilities; however, its sequential hybrid framework may increase the risk of overfitting. Therefore, this study evaluates the performance of Geographical-XGBoost (G-XGBoost), which directly integrates geographical weighting into the XGBoost algorithm.

This study aims to evaluate the performance of G-XGBoost compared with OLS, GWR, XGBoost, and GW-XGBoost across different levels of spatial heterogeneity and both linear and nonlinear relationship patterns. The evaluation was conducted through simulation studies involving four scenarios that combined low and high spatial heterogeneity with linear and nonlinear relationships. Subsequently, all models were applied to district- and city-level poverty data in Java Island for 2024 to determine the best-performing model. Model performance was assessed using the coefficient of determination (R^2) Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE), while model stability was evaluated based on the performance gap between the training and testing datasets.

The simulation results demonstrate that G-XGBoost achieved the most stable and consistent performance across all evaluated scenarios. In the linear scenarios, OLS and GWR produced competitive results; however, their performance deteriorated substantially under nonlinear conditions. In contrast, G-XGBoost consistently maintained high predictive accuracy across both linear and nonlinear relationships and under varying degrees of spatial heterogeneity. Furthermore, G-XGBoost exhibited the smallest gap between training and testing performance among all competing models, indicating superior generalization ability and a lower risk of overfitting than GW-XGBoost.

In the empirical study, G-XGBoost again achieved the best performance, with an R^2 value of 0.8115 and an RMSE of 1.6241, outperforming all competing models. The Local Variable Importance (LVI) analysis revealed that the relative importance of explanatory variables in predicting poverty varied across districts and cities in Java Island. Average years of schooling was among the variables that most frequently exhibited high variable importance values, while the percentage of the population employed in the agricultural sector, the percentage of households with semi-permanent flooring, minimum wage, health insurance coverage, and the percentage of the population with at least a senior high school education were also identified as variables with high variable importance in specific regions. These



findings demonstrate that G-XGBoost is capable of simultaneously accommodating nonlinear relationships and spatial heterogeneity while providing an initial reference for identifying explanatory variables with relatively high variable importance for further investigation according to the characteristics of each region.

Keywords: Geographical-XGBoost, Geographically Weighted Regression, poverty, spatial heterogeneity, XGBoost.

© 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.

EVALUASI KINERJA MODEL GEOGRAPHICAL-XGBOOST PADA DATA DENGAN TINGKAT HETEROGENITAS SPASIAL DAN POLA LINEARITAS BERBEDA.

ZAMRAH MUTMAINAH

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

- 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 Tesis:
Prof. Dr. Ir. Muhammad Nur Aidi, M.S.



Judul Tesis : Evaluasi Kinerja Model *Geographical-XGBoost* pada Data dengan Tingkat Heterogenitas Spasial dan Pola Linearitas Berbeda

Nama : Zamrah Mutmainah

NIM : M0501241066

Disetujui oleh

Pembimbing 1:

Dr. Ir. Budi Susetyo, M.S.



Pembimbing 2:

Dr. Septian Rahardianto, S.Stat., 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 Buono, M.Si., M.Kom.

NIP 19660702 199302 1 001



Tanggal Ujian:
15 Juli 2026

Tanggal Lulus:



PRAKATA

Puji syukur penulis panjatkan ke hadirat Allah SWT atas segala limpahan rahmat dan karunia-Nya, sehingga tesis yang berjudul “Evaluasi Kinerja Model *Geographical-XGBoost* pada Data dengan Tingkat Heterogenitas Spasial dan Pola Linearitas Berbeda” ini dapat diselesaikan dengan baik. Penyusunan tesis ini merupakan salah satu tahapan penting dalam pemenuhan syarat tugas akhir 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 Dr. Ir. Budi Susetyo, M.S. dan Bapak Dr. Septian Rahardiantoro, S.Stat., M.Si. selaku dosen pembimbing yang telah meluangkan waktu untuk memberikan arahan, ilmu, serta saran yang sangat berharga dalam penyempurnaan penelitian ini.
2. Lembaga Pengelola Dana Pendidikan (LPDP) atas kepercayaan dan dukungan finansial melalui beasiswa yang diberikan, sehingga penulis dapat menempuh studi dengan lancar.
3. Seluruh dosen dan Tenaga kependidikan Departemen Statistika IPB University atas bekal ilmu dan fasilitas yang sangat mendukung perjalanan akademik penulis.
4. Orang tua, keluarga besar, yang senantiasa memberikan doa, cinta, dan dukungan tanpa henti yang menjadi kekuatan bagi penulis.
5. 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, Juli 2026

Zamrah Mutmainah



- 

DAFTAR ISI

DAFTAR ISI	vi
DAFTAR TABEL	vii
DAFTAR GAMBAR	vii
I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Tujuan	3
1.4 Manfaat	4
II TINJAUAN PUSTAKA	5
2.1 <i>Machine Learning</i>	5
2.2 <i>Extreme Gradient Boosting (XGBoost)</i>	6
2.3 <i>Geographically Weighted Regression (GWR)</i>	7
2.4 <i>Geographically Weighted-XGBoost (GW-XGBoost)</i>	8
2.5 <i>Geographical - XGBoost (G-XGBoost)</i>	9
2.6 Breusch-Pagan	12
2.7 Uji Friedman	13
2.8 Uji Wilcoxon	13
2.9 <i>Locally Estimated Scatterplot Smoothing (LOESS)</i>	14
2.10 Evaluasi Model	14
III METODE	16
3.1 Sumber Data	16
3.1.1 Kajian Simulasi	16
3.1.2 Kajian Empiris	17
3.2 Prosedur Analisis	19
3.2.1 Kajian Simulasi	19
3.2.2 Kajian Empiris	19
IV HASIL DAN PEMBAHASAN	21
4.1. Kajian Simulasi	21
4.1.1 Desain Simulasi	21
4.1.2 Pemodelan	25
4.1.3. Perbandingan performa prediksi	26
4.1.4. Generalisasi Model	30
4.2 Kajian Empiris	34
4.2.1 Ekplorasi Data	34
4.2.2 Pemodelan	38
4.2.3 Perbandingan performa evaluasi model	39
4.2.4 <i>Local Variable Importance (LVI)</i>	41
V SIMPULAN DAN SARAN	46
5.1 Simpulan	46
5.2 Saran	46
DAFTAR PUSTAKA	48
LAMPIRAN	53
RIWAYAT HIDUP	67

DAFTAR TABEL

1	Nilai default <i>Hyperparameter tuning</i> yang digunakan	17
2	Data Empiris	18
3	Statistik deskriptif kinerja data pengujian (Rerata $\pm$ SD)	28
4	Uji Friedman untuk performa data uji	29
5	Perbandingan <i>post-hoc</i> Wilcoxon terhadap data uji	29
6	Statistik deskriptif kinerja selisih data latih dan uji (Rerata $\pm$ SD)	32
7	Uji Friedman untuk performa selisih data latih dan uji	33
8	Perbandingan <i>post-hoc</i> Wilcoxon terhadap performa gap	33
9	Uji Breusch-Pagan	36
10	Nilai <i>Hyperparameter tuning</i> yang digunakan	38
11	<i>Variable Importance</i> per Lokasi yang tertinggi	44

DAFTAR GAMBAR

1	Ilustrasi Alur GW-XGBoost	9
2	<i>Framework</i> metode G-XGBoost	12
3	Tahapan kajian simulasi	19
4	Tahapan kajian simulasi	20
5	Distribusi Permukaan Koefisien β : (a) Koefisien β_0 , (b) Koefisien β_1	21
6	Distribusi Permukaan Koefisien β : (a) Koefisien β_1^1 , (b) Koefisien β_2^2	22
7	Plot Hubungan Linear: (a) x_1 dan $y_{\text{skenario } 1}$, (b) x_2 dan $y_{\text{skenario } 1}$, (c) x_1 dan $y_{\text{skenario } 2}$, (d) x_2 dan $y_{\text{skenario } 2}$	23
8	Plot Hubungan Nonlinear: (a) x_1 dan $y_{\text{skenario } 3}$, (b) x_2 dan $y_{\text{skenario } 3}$, (c) x_1 dan $y_{\text{skenario } 4}$, (d) x_2 dan $y_{\text{skenario } 4}$	24
9	Boxplot Performa model G-XGBoost dengan variasi parameter α	26
10	_Toc233888199Boxplot Perbandingan Evaluasi Performa R^2 Model	26
11	Boxplot Perbandingan Evaluasi Performa RMSE Model	27
12	Boxplot Perbandingan Evaluasi Performa MAE Model	27
13	Boxplot <i>Gap</i> Generalisasi model berdasarkan metrik R^2	30
14	Boxplot <i>Gap</i> Generalisasi model berdasarkan metrik RMSE	31
15	Boxplot <i>Gap</i> Generalisasi model berdasarkan metrik MAE	31
16	Distribusi spasial tingkat kemiskinan di Pulau Jawa	35
17	Koefisien variasi peubah penjas	36
18	Pola hubungan peubah respon dan penjas	37
19	Perbandingan Evaluasi Performa R^2 Model	39
20	Perbandingan Evaluasi Performa RMSE Model	40
21	Perbandingan Evaluasi Performa MAE Model	40
22	<i>Variable importance</i> lokal	42



DAFTAR LAMPIRAN

1	Nilai <i>local variable importance</i> setiap lokasi pengamatan	54
2	Peta visualisasi <i>local variable importance</i> setiap lokasi pengamatan	59
3	Nilai R^2 Lokal	60
4	Hasil uji Friedman data uji	61
5	Hasil uji Wilcoxon data uji (R^2)	61
6	Hasil uji Wilcoxon data uji (RMSE)	62
7	Hasil uji Wilcoxon data uji (MAE)	63
8	Hasil uji Friedman performa gap	64
9	Hasil uji Wilcoxon data gap (R^2)	64
10	Hasil uji Wilcoxon data gap (RMSE)	65
11	Hasil uji Wilcoxon data gap (MAE)	66

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.



- 