

Prediksi Harga Rumah Jabodetabek Menggunakan Machine Learning dengan Fitur Geospasial dan Perspektif Pengembang Properti

Muhammad Alif Ramadhan^{1*)}, Maria Bestarina Laili², Shafa Aura Ayu³

^{1,3} Program Studi S1 Teknik Elektro, Universitas Singaperbangsa Karawang, Karawang, Indonesia

² Program Studi Teknik Elektro, Universitas Singaperbangsa Karawang, Karawang, Indonesia

*Penulis Korespondensi, e-mail: 2110631160014@student.unsika.ac.id

Received: 16/05/2026

Revised: 17/06/2026

Accepted: 17/06/2026

ABSTRAK

Penetapan harga rumah di kawasan Jabodetabek masih bergantung pada penilaian manual yang subjektif sehingga menimbulkan ketimpangan informasi antara penjual dan pembeli. Penelitian ini membandingkan empat algoritma pembelajaran mesin, yaitu regresi linear, hutan acak, peningkatan gradien berbasis histogram, dan ansambel bertumpuk, untuk memprediksi harga rumah berdasarkan 3.508 data listing dari sembilan kota dan kabupaten. Kontribusi utama penelitian adalah perancangan fitur dari perspektif pengembang properti, meliputi rasio koefisien dasar bangunan, kepastian legal sertifikat, tingkat daya listrik, skor fasilitas premium, dan indikator kawasan bermerek, yang dipadukan dengan fitur geospasial berupa jarak ke pusat ekonomi. Dua skenario fitur diuji untuk mengukur kontribusi fitur tersebut. Model peningkatan gradien menghasilkan kinerja terbaik dengan koefisien determinasi sebesar 0,929 dan galat persentase absolut rata-rata sebesar 18,84 persen. Penambahan fitur geospasial dan perspektif pengembang terbukti meningkatkan akurasi prediksi secara konsisten dan signifikan.

Kata Kunci: fitur geospasial, harga rumah, pembelajaran mesin, peningkatan gradien, perspektif pengembang, prediksi

ABSTRACT

Determining house prices in the Jabodetabek region still relies on subjective manual appraisal, creating information asymmetry between sellers and buyers. This study compares four machine learning algorithms, namely linear regression, random forest, histogram based gradient boosting, and stacking ensemble, to predict house prices based on 3,508 listing records from nine cities and regencies. The main contribution is the design of features from a property developer perspective, including building coverage ratio, legal certainty of certificates, electricity capacity, premium facility score, and branded township indicator, combined with geospatial features in the form of distance to economic centers. Two feature scenarios were tested to measure the contribution of these features. The gradient boosting model achieved the best performance with a coefficient of determination of 0.929 and a mean absolute percentage error of 18.84 percent. Adding geospatial and developer perspective features consistently and significantly improved prediction accuracy.

Keywords: Branded Township, Developer Perspective, Geospatial Features, Gradient Boosting, House Price, Machine Learning

1. PENDAHULUAN

Pasar properti residensial di kawasan Jakarta, Bogor, Depok, Tangerang, dan Bekasi merupakan salah satu yang paling dinamis dan heterogen di Indonesia, dengan rentang harga dari ratusan juta hingga puluhan miliar rupiah dalam satu wilayah aglomerasi. Penentuan harga yang masih mengandalkan penilaian agen secara manual rentan terhadap inkonsistensi dan ketimpangan informasi. Pendekatan pembelajaran

p-ISSN: 2356-0533; e-ISSN: 2355-9195



mesin menawarkan estimasi harga yang objektif, dapat direproduksi, dan terukur, serta terbukti lebih akurat dibanding metode penilaian tradisional [1], [2].

Metode ansambel berbasis pohon seperti hutan acak dan peningkatan gradien secara konsisten mengungguli model linear pada data properti karena mampu menangkap hubungan tidak linear dan interaksi antarfitur [3]. Hutan acak telah luas diterapkan untuk prediksi harga rumah dengan akurasi tinggi [4], dan studi perbandingan terbaru menempatkannya unggul atas model linear maupun pohon tunggal [5], sementara model ansambel bertumpuk yang menggabungkan beberapa pembelajar dilaporkan memberikan performa terbaik pada sejumlah studi mutakhir [6]. Pengembangan lanjutan seperti *XGBoost* [7] memperluas efisiensi metode tersebut, sementara kerangka interpretasi seperti *SHapley Additive exPlanations* [8] memungkinkan model yang kompleks tetap dapat dijelaskan dan diadopsi oleh penilai serta pemangku kepentingan properti [9].

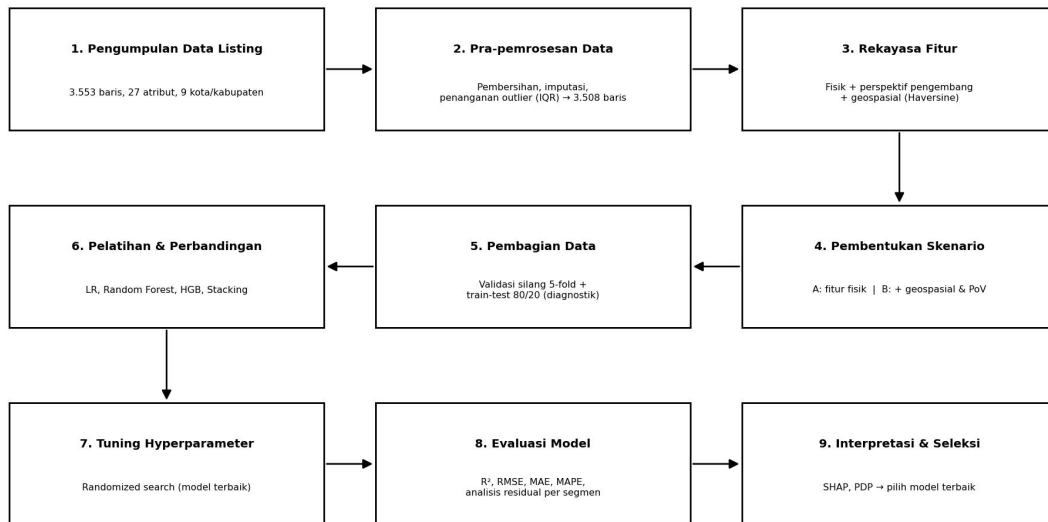
Sebagian besar studi prediksi harga properti berfokus pada fitur fisik bangunan dan belum memanfaatkan dua dimensi penting, yaitu konteks spasial lokasi terhadap pusat ekonomi serta atribut bernilai komersial dari sudut pandang pengembang properti seperti kepastian legal, segmentasi daya listrik, dan premium kawasan bermerek. Studi terbaru di kawasan Jabodetabek dan kota besar Indonesia lain telah menerapkan pembelajaran mesin untuk prediksi harga properti [10], [11], [12], namun atribut lokasi dan struktural masih menjadi penentu utama yang dominan [13]. Kebaruan penelitian ini terletak pada perancangan fitur perspektif pengembang dan pengujian kontribusinya secara terukur.

Tujuan penelitian dirumuskan dalam tiga pertanyaan. Pertama, algoritma mana yang paling akurat memprediksi harga rumah Jabodetabek. Kedua, seberapa besar kontribusi fitur geospasial dan perspektif pengembang terhadap akurasi. Ketiga, fitur apa yang paling menentukan harga beserta arah pengaruhnya. Kontribusi penelitian mencakup tolok ukur empat algoritma pada data properti Indonesia, rancangan fitur berbasis perspektif pengembang, bukti kuantitatif kontribusi fitur geospasial, serta penerapan interpretasi model untuk transparansi keputusan.

2. METODE PENELITIAN

Penelitian mengikuti kerangka kerja penambangan data lintas industri yang mencakup pemahaman data, pra-pemrosesan, rekayasa fitur, pemodelan, evaluasi, dan interpretasi. Seluruh implementasi menggunakan bahasa Python dengan pustaka *scikit-learn* [14]. Variabel target ditransformasi secara logaritmik menggunakan Persamaan (5) untuk menstabilkan distribusi harga yang menceng ke kanan. Alur lengkap tahapan penelitian, mulai dari pengumpulan data hingga pemilihan model terbaik, ditunjukkan pada Gambar 1.





Gambar 1: Diagram alir metodologi penelitian yang diusulkan

2.1 Dataset Penelitian

Data bersumber dari listing publik kawasan Jabodetabek yang terdiri atas 3.553 baris dan 27 atribut, mencakup harga, koordinat geografis, luas tanah dan bangunan, jumlah kamar, sertifikat, daya listrik, fasilitas, serta kondisi properti [19]. Setelah pra-pemrosesan diperoleh 3.508 listing bersih yang tersebar pada sembilan kota dan kabupaten.

2.2 Pra-pemrosesan Data

Tahapan pembersihan meliputi penghapusan kolom yang tidak prediktif dan bernilai konstan, penghapusan baris tanpa nilai target atau atribut fisik esensial, penanganan nilai hilang, serta penanganan pencilan. Atribut dengan nilai hilang tinggi yaitu orientasi bangunan sebesar 46 persen dan usia bangunan sebesar 41 persen dikeluarkan, sedangkan atribut dengan nilai hilang rendah diimputasi menggunakan median atau modus. Pencilan pada harga, luas tanah, dan luas bangunan ditangani dengan pembatasan berbasis rentang antarkuartil. Sebanyak 45 baris atau 1,3 persen dieliminasi pada proses ini.

2.3 Rekayasa Fitur Perspektif Pengembang

Selain fitur fisik mentah, dirancang fitur turunan yang merepresentasikan nilai komersial sebagaimana dipersepsi pengembang dan penilai properti. Tabel I merangkum fitur rekayasa beserta rasionalnya.

TABEL I: FITUR REKAYASA DAN RASIONAL PERSPEKTIF PENGEMBANG

Fitur	Definisi	Rasional
Rasio KDB	Luas bangunan dibagi luas tanah	Intensitas dan efisiensi pemanfaatan lahan
Skor legal	SHM = 3, HGB = 2, Girik = 1	Kepastian hukum menentukan daya jual
Daya listrik	Ekstraksi nilai watt dari atribut	Proxy segmentasi mewah dan menengah
Jumlah fasilitas	Hitung item pada teks fasilitas	Kelengkapan amenities kawasan
Skor fasilitas premium	Hitung amenities premium	Indikator pemosisian premium
Indikator keamanan	Ada atau tidak sistem keamanan	Proxy hunian dalam kawasan tertutup
Kawasan bermerek	Kecocokan nama pengembang ternama	Premium merek pengembang
Jarak ke pusat kota	Jarak Haversine ke pusat ekonomi	Aksesibilitas ke pusat kegiatan

Agar penelitian dapat direplikasi, prosedur pembentukan fitur utama dinyatakan secara formal. Rasio koefisien dasar bangunan (KDB) sebagai proksi intensitas pemanfaatan lahan didefinisikan pada Persamaan (1). Skor legal sertifikat ditetapkan secara ordinal pada Persamaan (2). Skor fasilitas premium dihitung sebagai jumlah amenities premium yang muncul pada teks fasilitas melalui Persamaan (3), dengan K adalah

himpunan kata kunci amenitas premium (kolam renang, keamanan, klub, gym, jogging track, dan sejenisnya). Jarak ke pusat ekonomi dihitung menggunakan formula Haversine pada Persamaan (4), dengan ϕ adalah lintang, λ adalah bujur, dan $R = 6.371$ km adalah jari-jari bumi. Harga per meter persegi hanya digunakan untuk eksplorasi dan segmentasi, tidak sebagai prediktor, guna menghindari kebocoran target.

$$KDB = \text{Luas Bangunan} / \text{Luas Tanah} \quad (1)$$

$$\text{Skor legal} = 3 \text{ (SHM)}; 2 \text{ (HGB)}; 1 \text{ (Girik/PPJB)}; 0 \text{ (lainnya)} \quad (2)$$

$$\text{Skor fasilitas premium} = \sum I(k \in \text{fasilitas}), k \in K \quad (3)$$

$$d = 2R \cdot \arcsin(\sqrt{(\sin^2(\Delta\phi/2) + \cos\phi_1 \cos\phi_2 \sin^2(\Delta\lambda/2))}) \quad (4)$$

2.4 Analisis Multivariat

Tiga analisis dilakukan untuk memahami struktur antarfitur, yaitu uji multikolinearitas melalui faktor inflasi varians, analisis komponen utama untuk reduksi dimensi dan interpretasi arah fitur, serta segmentasi pasar menggunakan K-Means dengan jumlah kluster optimal ditentukan oleh skor siluet.

2.5 Pemodelan dan Skenario Pengujian

Empat algoritma dibandingkan, yaitu regresi linear sebagai garis dasar, hutan acak [15], peningkatan gradien berbasis histogram [16], dan ansambel bertumpuk. Arsitektur ansambel bertumpuk [17] menggunakan hutan acak, peningkatan gradien, dan regresi Ridge sebagai pembelajar dasar dengan regresi Ridge sebagai pembelajar meta. Peningkatan gradien berbasis histogram dipilih sebagai representasi keluarga peningkatan gradien karena tersedia secara native pada pustaka yang digunakan dan sepenuhnya dapat direproduksi.

Dua skenario fitur dirancang untuk menjawab pertanyaan kontribusi fitur. Skenario A hanya menggunakan fitur fisik bangunan, sedangkan Skenario B menambahkan fitur geospasial, fitur perspektif pengembang, dan penyandian lokasi kategorikal.

2.6 Validasi dan Penyetelan Hyperparameter

Untuk menjamin perbandingan yang adil, seluruh model dievaluasi dengan prosedur identik berupa validasi silang lima lipatan (5-fold cross validation) yang diacak. Seluruh tahapan pra-pemrosesan, penskalaan, dan penyandian dibungkus dalam satu alur (pipeline) yang dilatih hanya pada data latih di setiap lipatan, sehingga tidak terjadi kebocoran data. Selain itu, pembagian data latih dan uji sebesar 80:20 digunakan secara terpisah untuk diagnostik residual dan interpretasi model. Validasi berbasis wilayah tidak digunakan agar perbandingan antaralgoritma tetap setara, dan disarankan sebagai pengembangan lanjutan.

Hyperparameter model terbaik dioptimasi menggunakan pencarian acak (randomized search) dengan validasi silang atas ruang pencarian pada Tabel V, sebagaimana ditekankan studi terdahulu untuk memastikan perbandingan yang adil dan reproduktibel [18]. Konfigurasi terpilih kemudian dievaluasi ulang dengan validasi silang lima lipatan yang sama.

TABEL II: RUANG PENCARIAN DAN NILAI TERPILIH HYPERPARAMETER MODEL TERBAIK

Hyperparameter	Ruang pencarian	Nilai terpilih
learning_rate	{0,03; 0,05; 0,08; 0,1}	0,08
max_iter	{200; 300; 400}	300
max_leaf_nodes	{15; 31; 63}	31
min_samples_leaf	{10; 20; 40}	20
l2_regularization	{0; 0,1; 1,0}	0

2.7 Metrik Evaluasi

Kinerja diukur dengan koefisien determinasi (R^2) pada skala log-harga melalui Persamaan (5), serta akar rata-rata galat kuadrat (RMSE), rata-rata galat absolut (MAE), dan galat persentase absolut rata-rata



(MAPE) yang dihitung pada skala harga asli setelah transformasi balik melalui Persamaan (6) sampai (8). Pada persamaan tersebut, y_i adalah harga aktual, $\hat{y}_i$ adalah harga prediksi, $\bar{y}$ adalah rata-rata harga, dan n adalah jumlah data. MAPE dipilih karena mudah ditafsirkan sebagai rata-rata persentase deviasi prediksi.

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} \quad (5)$$

$$RMSE = \sqrt{\frac{1}{n} \sum (y_i - \hat{y}_i)^2} \quad (6)$$

$$MAE = \frac{1}{n} \sum |y_i - \hat{y}_i| \quad (7)$$

$$MAPE = \frac{100\%}{n} \sum |(y_i - \hat{y}_i) / y_i| \quad (8)$$

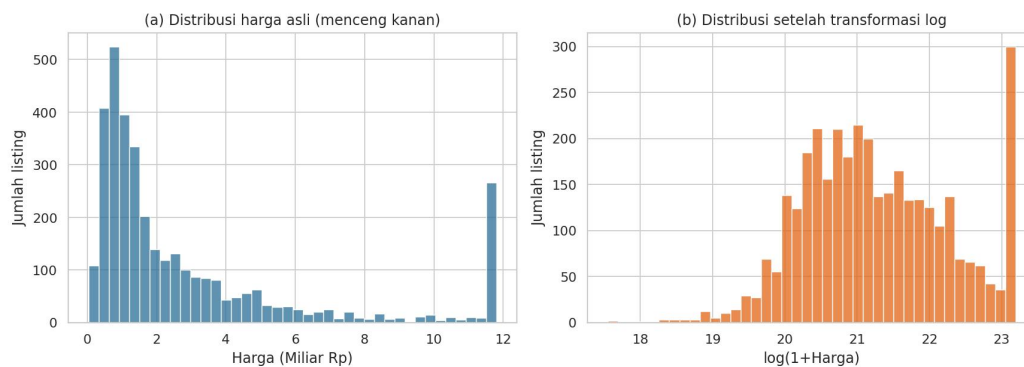
2.8 Interpretabilitas Model

Model terbaik diinterpretasi menggunakan nilai SHapley untuk memperoleh kepentingan fitur global, arah pengaruh tiap fitur melalui diagram sebar ringkas, serta penjelasan lokal pada satu listing [8]. Diagram ketergantungan parsial digunakan untuk memvisualisasikan bentuk pengaruh marginal fitur kunci terhadap harga.

3. HASIL DAN PEMBAHASAN

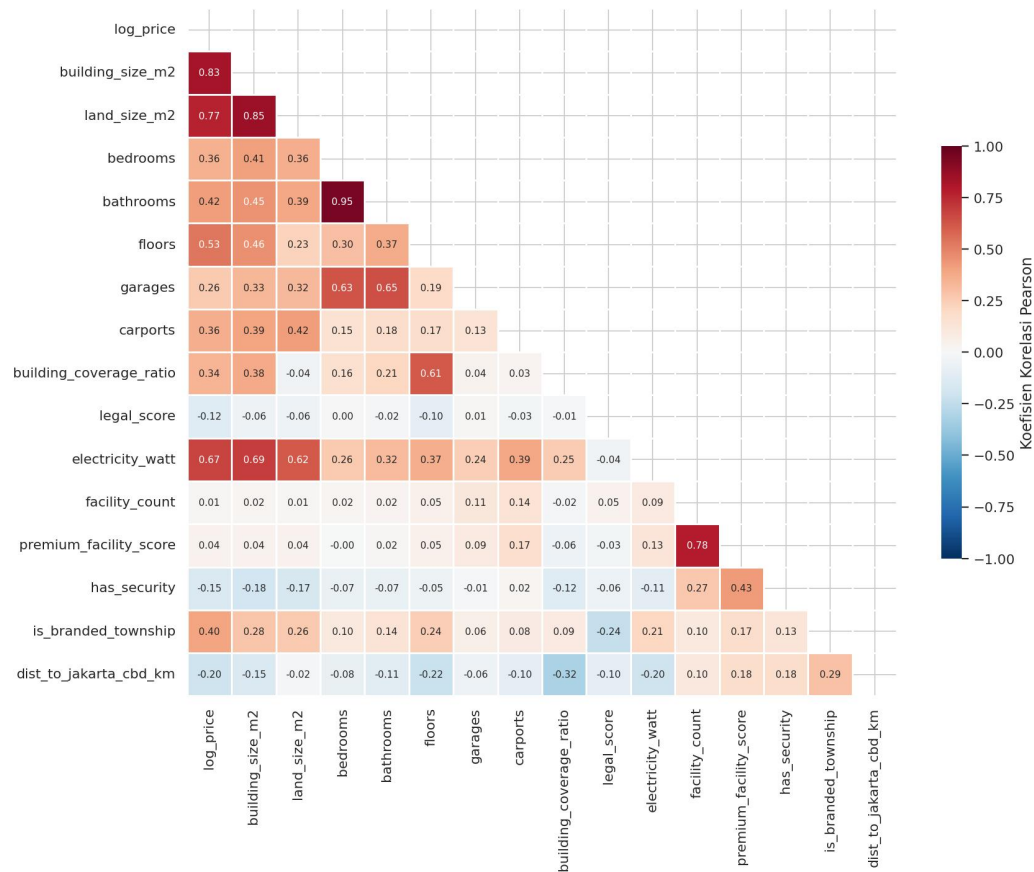
3.1 Statistik Deskriptif dan Eksplorasi Data

Harga rumah memiliki median sebesar Rp 1,5 miliar dengan distribusi sangat menceng ke kanan. Transformasi logaritmik berhasil menormalkan distribusi target sebagaimana ditunjukkan Gambar 2, sehingga lebih sesuai untuk pemodelan regresi.

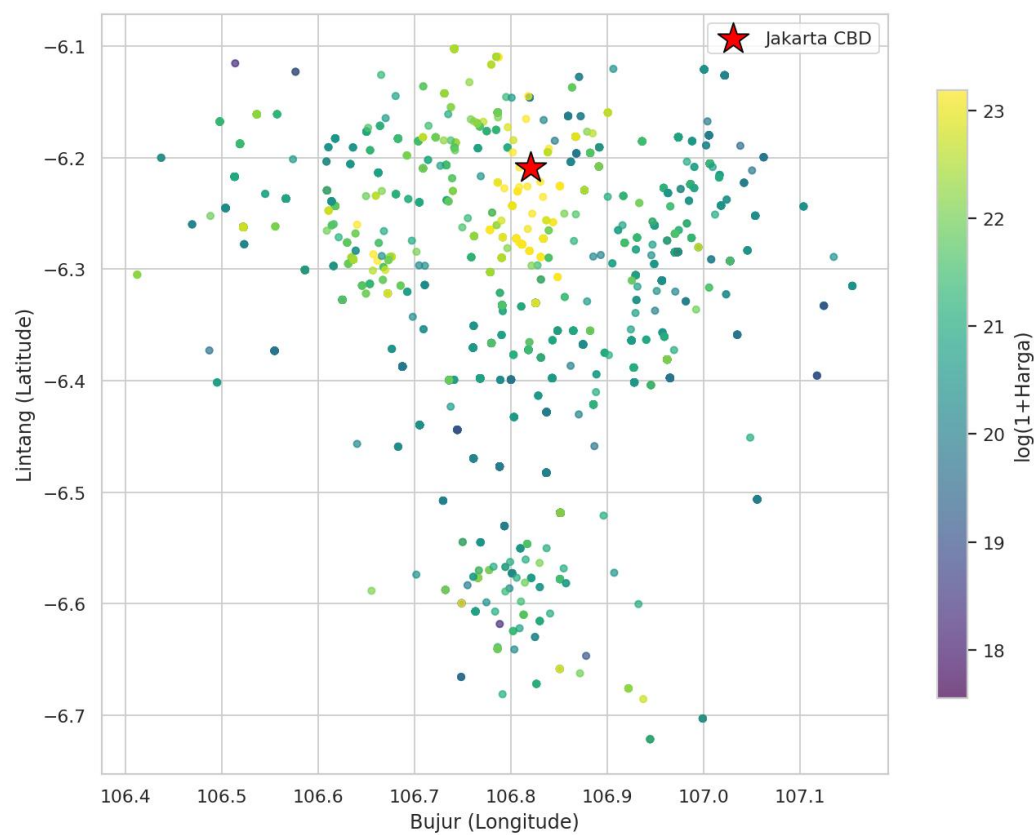


Gambar 2: Distribusi harga sebelum dan sesudah transformasi logaritmik

Matriks korelasi pada Gambar 3 memperlihatkan luas bangunan dan luas tanah sebagai fitur paling berkorelasi dengan harga. Secara spasial, harga menurun seiring bertambahnya jarak dari pusat bisnis Jakarta sebagaimana terlihat pada Gambar 4, yang menegaskan pentingnya dimensi geospasial.

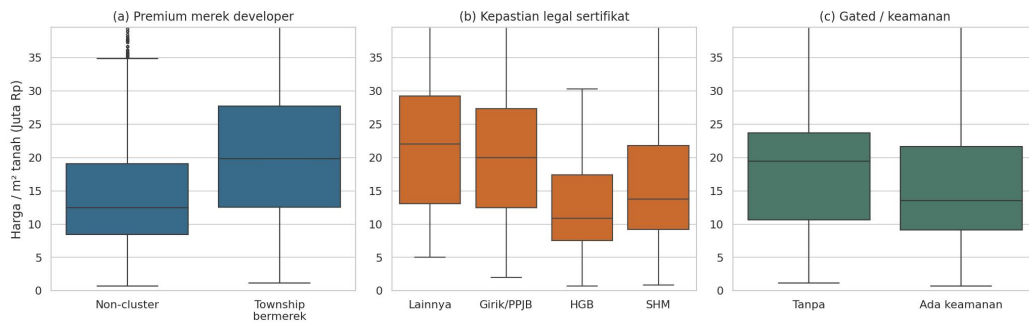


Gambar 3: Matriks korelasi fitur fisik, geospasial, dan perspektif pengembang



Gambar 4: Sebaran spasial harga; titik lebih terang menandakan harga lebih tinggi mendekati pusat kota

Analisis univariat fitur perspektif pengembang pada Gambar 5 menunjukkan premium yang konsisten untuk kawasan bermerek, sedangkan asosiasi sertifikat dan keamanan bersifat tidak monoton akibat efek pembaur lokasi. Temuan ini justru memotivasi pendekatan multivariat dan tidak linear.



Gambar 5: Asosiasi univariat fitur perspektif pengembang dengan harga per meter persegi tanah

3.2 Analisis Multivariat

Uji faktor inflasi varians pada Tabel III mengungkap multikolinearitas tinggi pada gugus fitur ukuran dan kamar dengan nilai lebih besar dari delapan. Temuan ini menjadi justifikasi pemilihan model ansambel berbasis pohon yang robust terhadap multikolinearitas, sekaligus menjelaskan kinerja buruk regresi linear pada skala harga asli. Analisis komponen utama menunjukkan dibutuhkan sembilan komponen untuk menjelaskan 90 persen varians, yang mengindikasikan informasi tersebar pada banyak dimensi.

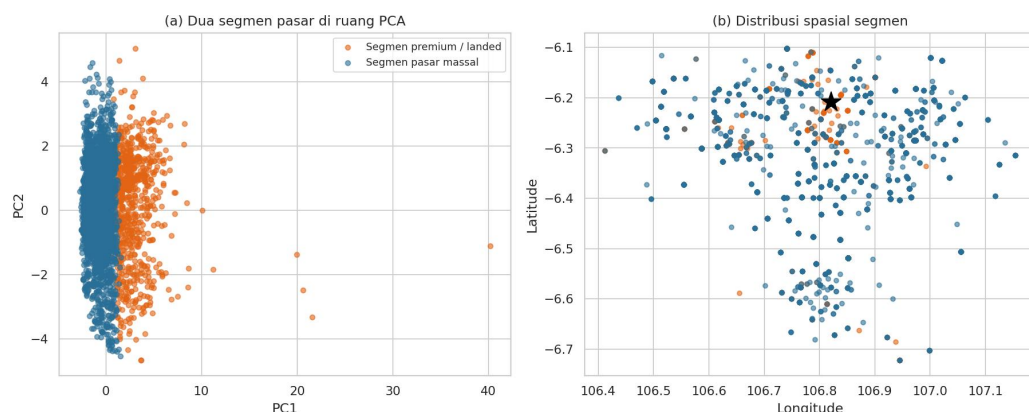
TABEL III: FAKTOR INFLASI VARIANS FITUR TERPILIH

Fitur	Nilai	Interpretasi
Jumlah kamar mandi	11,09	Multikolinearitas tinggi
Luas bangunan	10,35	Multikolinearitas tinggi
Jumlah kamar tidur	10,11	Multikolinearitas tinggi
Luas tanah	8,84	Multikolinearitas sedang tinggi
Rasio KDB	3,50	Dapat diterima
Jarak ke pusat kota	1,38	Rendah dan independen
Skor legal	1,08	Rendah dan independen

Segmentasi K-Means menghasilkan dua segmen optimal yang merepresentasikan pasar properti terpolarisasi, yaitu segmen hunian mewah dan segmen pasar massal, sebagaimana dirangkum pada Tabel IV dan divisualisasikan pada Gambar 6.

TABEL IV: PROFIL SEGMENT PASAR HASIL K-MEANS

Segmen	Jumlah	Harga med.	Bangunan	Tanah	Jarak	Bermerek
Premium	743	Rp 7,6 M	400 m2	340 m2	20,8 km	55%
Massal	2.765	Rp 1,2 M	90 m2	90 m2	22,6 km	22%



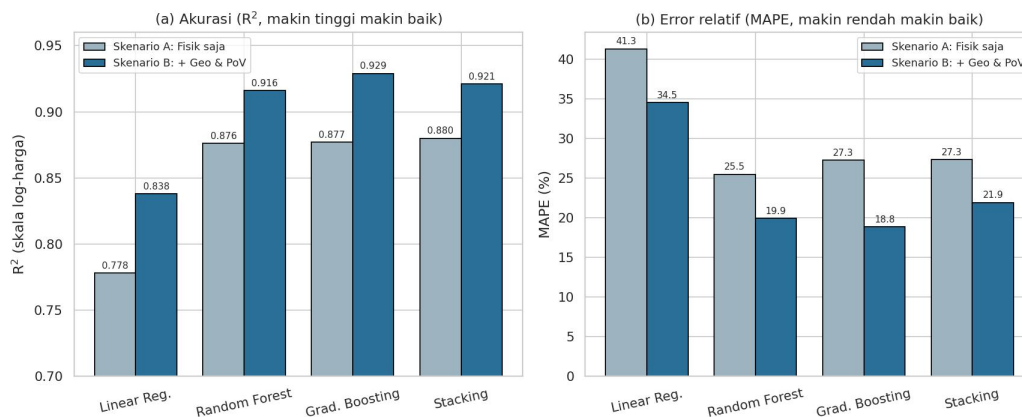
Gambar 6: Segmentasi pasar properti Jabodetabek pada ruang komponen utama dan ruang spasial

3.3 Perbandingan Kinerja Model

Tabel V merangkum kinerja keempat algoritma pada kedua skenario. Tiga temuan utama diperoleh. Pertama, penambahan fitur geospasial dan perspektif pengembang pada Skenario B meningkatkan akurasi seluruh model secara konsisten, dengan galat persentase model terbaik turun dari 27,28 persen menjadi 18,84 persen. Kedua, model ansambel berbasis pohon jauh mengungguli regresi linear. Ketiga, peningkatan gradien menjadi model terbaik dan sedikit di atas ansambel bertumpuk, yang mengindikasikan bahwa pembelajar dasar yang berkorelasi membatasi keuntungan penumpukan pada kasus ini.

TABEL V: PERBANDINGAN KINERJA MODEL DENGAN VALIDASI SILANG LIMA LIPATAN

Skenario dan Model	R ² (log)	RMSE (M)	MAE (M)	MAPE (%)
A. Regresi Linear	0,778	3,253	1,242	41,31
A. Hutan Acak	0,876	1,315	0,644	25,48
A. Peningkatan Gradien	0,877	1,332	0,692	27,28
A. Ansambel Bertumpuk	0,880	1,337	0,695	27,33
B. Regresi Linear	0,838	2,548	1,023	34,55
B. Hutan Acak	0,916	1,000	0,470	19,90
B. Ansambel Bertumpuk	0,921	1,331	0,630	21,92
B. Peningkatan Gradien (terbaik)	0,929	0,920	0,451	18,84



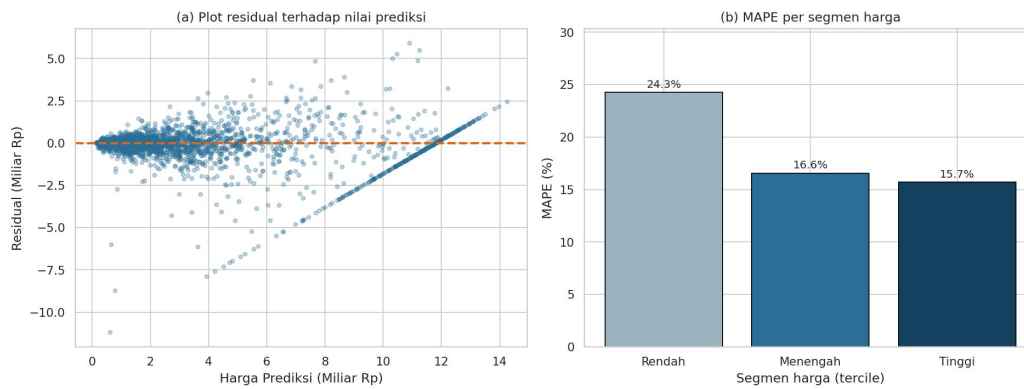
Gambar 7: Perbandingan empat algoritma pada dua skenario fitur

3.4 Penyetelan Hyperparameter dan Analisis Residual

Pencarian acak hyperparameter atas ruang pada Tabel V mengonfirmasi bahwa konfigurasi peningkatan gradien dengan learning rate 0,08 dan 300 iterasi merupakan yang terbaik, dengan R^2 sebesar 0,929 dan MAPE sebesar 18,84 persen pada validasi silang lima lipatan. Plot residual pada Gambar 8(a) memperlihatkan sebaran galat yang relatif simetris di sekitar nol, meskipun terdapat sedikit pelebaran varians pada properti bernilai sangat tinggi, suatu pola heteroskedastisitas yang lazim pada data properti.

Analisis galat per segmen harga pada Gambar 8(b) menunjukkan model paling akurat pada properti harga tinggi dengan MAPE 15,7 persen dan menengah 16,6 persen, tetapi galat relatif meningkat menjadi 24,3 persen pada segmen harga rendah. Pola ini mengindikasikan bias model pada rumah berharga rendah, sehingga estimasi pada segmen tersebut perlu disikapi dengan hati-hati.

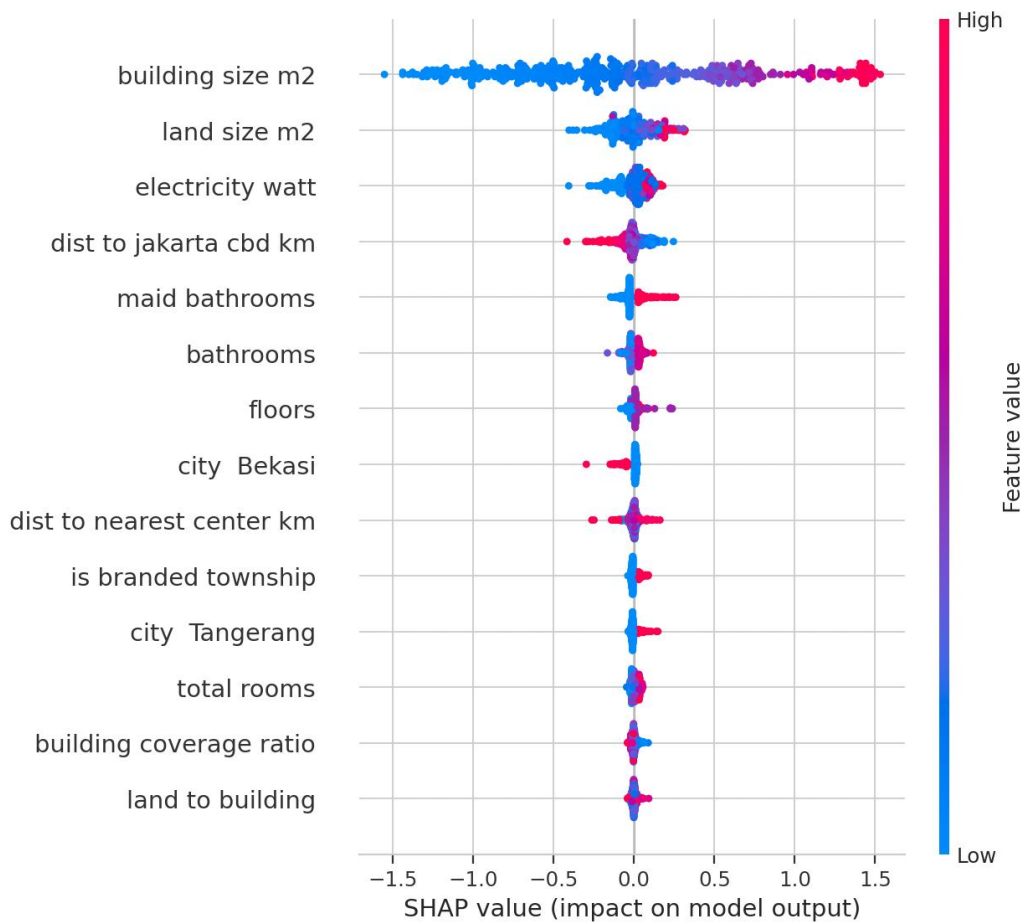




Gambar 8: Plot residual dan galat persentase per segmen harga model terbaik

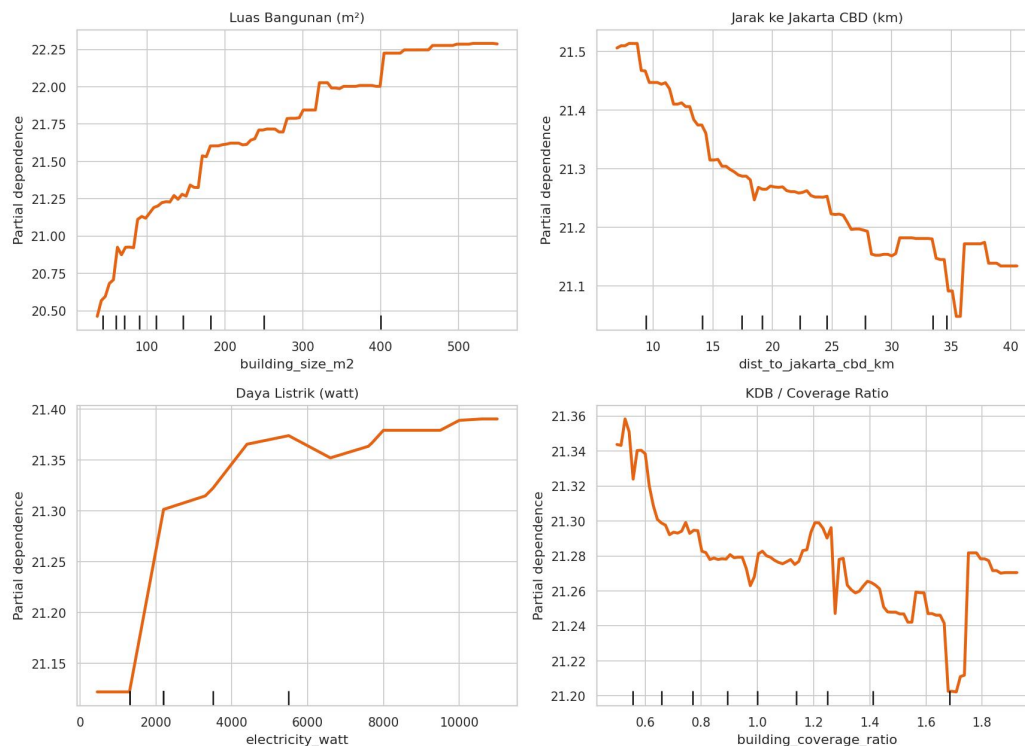
3.5 Interpretasi Model

Analisis nilai SHapley pada Gambar 9 mengidentifikasi luas bangunan sebagai penggerak harga paling dominan, diikuti luas tanah, daya listrik, dan jarak ke pusat kota. Penting dicatat bahwa empat dari sepuluh fitur teratas merupakan fitur rekayasa geospasial dan perspektif pengembang, yaitu daya listrik, jarak ke pusat kota, jarak ke pusat terdekat, dan kawasan bermerek. Hal ini membuktikan nilai tambah rancangan fitur yang diusulkan.



Gambar 9: Diagram ringkas nilai SHapley untuk arah dan besar pengaruh fitur terhadap log-harga

Diagram ketergantungan parsial pada Gambar 10 mengonfirmasi bentuk hubungan marginal. Harga meningkat tajam terhadap luas bangunan lalu melandai, dan menurun secara monoton terhadap jarak ke pusat kota. Pola tidak linear ini menjelaskan keunggulan model berbasis pohon dibanding pendekatan linear.



Gambar 10: Ketergantungan parsial fitur kunci terhadap log-harga

3.6 Implikasi Praktis

Hasil penelitian memberikan beberapa implikasi praktis. Model dapat berfungsi sebagai alat estimasi harga objektif dengan deviasi rata-rata di bawah 19 persen, dengan catatan akurasi pada segmen harga rendah masih perlu ditingkatkan. Faktor legal dan daya listrik terbukti bernilai sehingga relevan dalam strategi pemosisian produk. Premium kawasan bermerek terkuantifikasi sehingga memberikan dasar empiris bagi keputusan penjenamaan kawasan. Kerangka interpretasi memastikan estimasi dapat dijelaskan kepada pemangku kepentingan.

3. KESIMPULAN DAN SARAN

Penelitian ini membandingkan empat algoritma pembelajaran mesin untuk prediksi harga rumah Jabodetabek dan menunjukkan bahwa peningkatan gradien berbasis histogram memberikan kinerja terbaik dengan koefisien determinasi sebesar 0,929 dan galat persentase absolut rata-rata sebesar 18,84 persen. Penambahan fitur geospasial dan fitur berbasis perspektif pengembang terbukti meningkatkan akurasi secara konsisten dan signifikan, dengan analisis nilai SHapley mengonfirmasi kontribusi nyata fitur tersebut.

Optimasi hyperparameter melalui pencarian acak menegaskan kestabilan konfigurasi model terbaik, sementara analisis residual mengungkap galat yang lebih tinggi pada segmen harga rendah. Keterbatasan penelitian mencakup cakupan data dari satu sumber listing, ketiadaan variabel makroekonomi temporal, serta belum diterapkannya validasi berbasis wilayah. Saran untuk penelitian lanjutan adalah mengintegrasikan data titik minat eksternal dan harga historis, menerapkan validasi berbasis wilayah, memperbaiki akurasi pada segmen harga rendah, serta menguji algoritma peningkatan gradien lain seperti *XGBoost* dan *LightGBM*.

DAFTAR PUSTAKA

- [1] W. K. O. Ho, B. S. Tang, and S. W. Wong, "Predicting property prices with machine learning algorithms," *Journal of Property Research*, vol. 38, no. 1, pp. 48–70, 2021.



- [2] R. A. Mora-García, M. F. Céspedes-López, and V. R. Pérez-Sánchez, "Housing price prediction using machine learning algorithms in COVID-19 times," *Land*, vol. 11, no. 11, 2022.
- [3] A. Soltani, M. Heydari, F. Aghaei, and C. J. Pettit, "Housing price prediction incorporating spatio-temporal dependency into machine learning algorithms," *Cities*, vol. 131, 2022.
- [4] A. B. Adetunji, O. N. Akande, F. A. Ajala, O. Oyewo, Y. F. Akande, and G. Oluwadara, "House price prediction using random forest machine learning technique," *Procedia Computer Science*, vol. 199, pp. 806–813, 2022.
- [5] Y. Jing, "Comparing machine learning models for house price prediction: linear regression, decision tree, and random forest," *ITM Web of Conferences*, vol. 73, 2025.
- [6] M. Dsilva et al., "House price prediction: a comparative study of machine learning ensemble techniques and deep learning model," in *Proc. IEEE Int. Conf. for Women in Innovation, Technology & Entrepreneurship (ICWITE)*, 2025.
- [7] T. Chen and C. Guestrin, "XGBoost: a scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, San Francisco, CA, USA, 2016, pp. 785–794.
- [8] S. M. Lundberg and S. I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems*, vol. 30, 2017, pp. 4765–4774.
- [9] M. C. Iban, "An explainable model for the mass appraisal of residences: the application of tree-based machine learning algorithms and interpretation of value determinants," *Habitat International*, vol. 128, 2022.
- [10] H. Santoso et al., "Enhancing geospatial house price prediction in Greater Jakarta using XGBoost and ResNet18 feature fusion," *JUITA: Jurnal Informatika*, 2026.
- [11] A. M. Saputro et al., "Predicting property prices using MLR, gradient boosting, and random forest: a case study in South Tangerang, Indonesia," in *Proc. 8th Int. Conf. New Media Studies (CONMEDIA)*, 2025.
- [12] T. Wiradinata et al., "Post-pandemic analysis of house price prediction in Surabaya: a machine learning approach," *Journal of Southwest Jiaotong University*, vol. 57, no. 5, 2022.
- [13] N. H. Zulkifley et al., "House price prediction using a machine learning model: a survey of literature," *International Journal of Modern Education and Computer Science*, vol. 12, no. 6, pp. 46–54, 2020.
- [14] F. Pedregosa et al., "Scikit-learn: machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [15] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [16] J. H. Friedman, "Greedy function approximation: a gradient boosting machine," *The Annals of Statistics*, vol. 29, no. 5, pp. 1189–1232, 2001.
- [17] D. H. Wolpert, "Stacked generalization," *Neural Networks*, vol. 5, no. 2, pp. 241–259, 1992.
- [18] D. Andrade-Girón et al., "Intelligent feature selection ensemble model for price prediction in real estate markets," *Informatics*, vol. 12, 2025.
- [19] N. Barizki, "Daftar harga rumah Jabodetabek," *Kaggle*, 2022. [Online]. Available: <https://www.kaggle.com/datasets/nafisbarizki/daftar-harga-rumah-jabodetabek>. [Accessed: Jun. 11, 2026].