

Prediksi Harga Rumah Berbasis *Machine Learning* dengan *Explainable AI* untuk Interpretabilitas Faktor Penentu

Hadijah^{1*}, Wiwin Handoko², Rizty Maulida Badri³

¹ Fakultas Matematika dan Ilmu Pengetahuan Alam, Statistika, Universitas Negeri Medan, Medan, Indonesia

² Fakultas Matematika dan Ilmu Pengetahuan Alam, Ilmu Komputer, Universitas Negeri Medan, Medan, Indonesia

³ Fakultas Matematika dan Ilmu Pengetahuan Alam, Fisika, Universitas Negeri Medan, Medan, Indonesia

Email: ^{1*}hadijahh@unimed.ac.id, ²wiwinhandoko@unimed.ac.id, ³riztymaulida@unimed.ac.id

Email Penulis Korespondensi: hadijahh@unimed.ac.id

Abstrak— Harga rumah ditentukan oleh banyak faktor yang saling memengaruhi, sehingga diperlukan metode prediksi yang tidak hanya akurat tetapi juga mudah dipahami oleh pelaku bisnis properti, investor, dan pengambil kebijakan. Penelitian ini bertujuan membangun model prediksi harga rumah menggunakan pendekatan *machine learning* yang dipadukan dengan *Explainable Artificial Intelligence (XAI)* agar hasil prediksi lebih transparan dan dapat diinterpretasikan dengan baik. Data yang digunakan berasal dari properti nyata dengan beberapa variabel utama, seperti luas bangunan, luas tanah, jumlah kamar tidur, jumlah kamar mandi, serta kapasitas garasi. Dalam penelitian ini, empat algoritma *machine learning* diuji dan dibandingkan, yaitu *Linear Regression*, *Random Forest*, *XGBoost*, dan *Gradient Boosting*. Kinerja masing-masing model dievaluasi menggunakan beberapa metrik, meliputi *Mean Absolute Error (MAE)*, *Root Mean Squared Error (RMSE)*, koefisien determinasi (R^2), dan *Mean Absolute Percentage Error (MAPE)*. Berdasarkan hasil pengujian, algoritma *Random Forest* memberikan performa terbaik dengan nilai R^2 sebesar 0,7636, MAE mencapai Rp1,305 miliar, RMSE sebesar Rp2,037 miliar, dan MAPE sebesar 25,84%. Selanjutnya, model terbaik dianalisis menggunakan metode *SHapley Additive exPlanations (SHAP)* untuk melihat interpretasi model secara global maupun lokal, serta *Local Interpretable Model-agnostic Explanations (LIME)* untuk menjelaskan hasil prediksi pada tiap data secara individual. Hasil analisis menunjukkan bahwa luas bangunan dan luas tanah menjadi faktor yang paling berpengaruh terhadap harga rumah. Pendekatan yang diterapkan dalam penelitian ini terbukti mampu meningkatkan keterbukaan model prediksi sehingga hasilnya lebih mudah dipahami dan lebih dipercaya oleh pengguna.

Kata Kunci: Machine Learning, Explainable AI, Prediksi Harga Rumah, SHAP, LIME

Abstract—House prices are determined by numerous interrelated factors, making it essential to develop prediction methods that are not only accurate but also interpretable by property business practitioners, investors, and policymakers. This study aims to construct a house price prediction model using a machine learning approach integrated with Explainable Artificial Intelligence (XAI) to produce predictions that are more transparent and comprehensibly interpretable. The data used in this study were derived from real property listings, incorporating several key variables including building area, land area, number of bedrooms, number of bathrooms, and garage capacity. Four machine learning algorithms were evaluated and compared, namely Linear Regression, Random Forest, XGBoost, and Gradient Boosting. The performance of each model was assessed using multiple evaluation metrics, comprising Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), coefficient of determination (R^2), and Mean Absolute Percentage Error (MAPE). Experimental results demonstrate that the Random Forest algorithm achieved the best performance, yielding an R^2 value of 0.7636, MAE of IDR 1.305 billion, RMSE of IDR 2.037 billion, and MAPE of 25.84%. The best-performing model was subsequently analyzed using SHapley Additive exPlanations (SHAP) to provide both global and local model interpretability, as well as Local Interpretable Model-agnostic Explanations (LIME) to explain individual predictions at the instance level. The analysis reveals that building area and land area are the most influential factors in determining house prices. The proposed approach demonstrates a measurable improvement in model transparency, rendering prediction outcomes more comprehensible and trustworthy for end users.

Keywords: Machine Learning, Explainable AI, House Price Prediction, SHAP, LIME

1. PENDAHULUAN

Sektor properti menjadi salah satu penggerak penting dalam pertumbuhan ekonomi Indonesia dan terus menunjukkan perkembangan yang cukup pesat setiap tahunnya. Harga rumah sebagai indikator utama dalam pasar properti dipengaruhi oleh banyak aspek yang saling berhubungan, baik dari sisi karakteristik bangunan seperti luas bangunan, luas tanah, jumlah kamar tidur, dan kamar mandi, maupun faktor eksternal seperti lokasi, akses terhadap fasilitas umum, serta kondisi ekonomi secara umum [1]. Banyaknya variabel yang memengaruhi harga rumah membuat proses prediksi menjadi cukup kompleks dan sulit diselesaikan hanya dengan pendekatan statistik konvensional yang cenderung terbatas pada hubungan linier antarvariabel [2]. Selain itu, laju urbanisasi yang terus meningkat di berbagai kota besar di Indonesia menyebabkan kebutuhan hunian semakin tinggi dibandingkan ketersediaannya. Kondisi ini menjadikan sistem prediksi harga rumah yang akurat semakin dibutuhkan sebagai acuan bagi masyarakat, investor, pengembang, maupun pemerintah [3].

Kesalahan dalam menentukan estimasi harga properti dapat menimbulkan berbagai dampak bagi banyak pihak. Dari sisi konsumen, harga rumah yang terlalu tinggi atau terlalu rendah dapat menyebabkan kerugian finansial. Bagi investor dan pengembang, prediksi harga yang tidak tepat berpotensi memengaruhi keputusan investasi dan perencanaan bisnis. Sementara itu, pemerintah dan pembuat kebijakan memerlukan model prediksi yang andal untuk mendukung penyusunan kebijakan di sektor perumahan dan tata kota [4]. Oleh sebab itu, pengembangan sistem prediksi harga rumah yang lebih akurat menjadi kebutuhan yang semakin penting di era digital saat ini.

Perkembangan teknologi *machine learning* dalam beberapa tahun terakhir memberikan solusi yang lebih efektif dibandingkan metode statistik tradisional, terutama dalam menangani hubungan data yang kompleks dan non-linier pada harga properti. Berbagai algoritma seperti *Random Forest*, *XGBoost*, dan *Gradient Boosting* telah banyak digunakan dan terbukti menghasilkan performa prediksi yang lebih baik dibandingkan regresi linier [5]. Penelitian yang dilakukan oleh Warjiyono dan rekan-rekannya [6], misalnya, memanfaatkan algoritma *Random Forest* untuk memprediksi harga rumah menggunakan variabel luas tanah, luas bangunan, jumlah kamar tidur, kamar mandi, dan garasi, dengan tingkat akurasi mencapai 75,10%. Penelitian lain oleh Fitri [2] menunjukkan bahwa metode berbasis *ensemble* seperti *Random Forest* dan *Gradient Boosted Trees* memiliki kemampuan yang lebih baik dalam mengolah data properti dibandingkan *Linear Regression*, terutama pada data dengan distribusi yang tidak normal. Hasil serupa juga ditemukan oleh Wulan Sari dan Prabowo [7] yang membandingkan *Random Forest*, *Gradient Boosting*, dan *XGBoost*. Penelitian tersebut menunjukkan bahwa model *ensemble* mampu memberikan hasil prediksi yang lebih stabil dan konsisten dibandingkan pendekatan regresi konvensional.

Walaupun demikian, sebagian besar model *machine learning* yang digunakan dalam prediksi harga rumah masih memiliki kelemahan dari sisi interpretasi. Banyak model bekerja layaknya *black box*, yaitu mampu menghasilkan prediksi yang akurat tetapi sulit menjelaskan alasan di balik hasil prediksi tersebut. Kondisi ini menjadi perhatian penting karena rendahnya transparansi model dapat mengurangi tingkat kepercayaan pengguna, terutama pada bidang yang berkaitan dengan keputusan bernilai besar seperti properti [8]. Sejumlah penelitian sebelumnya juga menegaskan bahwa kurangnya kemampuan model dalam memberikan penjelasan menjadi salah satu hambatan utama dalam penerapan teknologi kecerdasan buatan secara lebih luas di sektor properti [4].

Sebagai solusi atas permasalahan tersebut, konsep *Explainable Artificial Intelligence (XAI)* mulai berkembang untuk meningkatkan transparansi dan interpretabilitas model *machine learning*. Dua metode yang paling sering digunakan adalah *SHapley Additive exPlanations (SHAP)* dan *Local Interpretable Model-agnostic Explanations (LIME)*. *SHAP* memanfaatkan konsep nilai *Shapley* dari teori permainan untuk menghitung kontribusi setiap fitur terhadap hasil prediksi, sehingga mampu memberikan penjelasan baik secara global maupun lokal [9]. Di sisi lain, *LIME* bekerja dengan membangun model sederhana di sekitar data tertentu agar proses pengambilan keputusan model dapat dipahami dengan lebih mudah [9].

Beberapa penelitian internasional telah menerapkan pendekatan *XAI* dalam prediksi harga properti dan menunjukkan hasil yang cukup baik. Ho [10], misalnya, mengembangkan model prediksi harga properti berbasis *XGBoost* yang dipadukan dengan *SHAP* menggunakan data transaksi properti di Hong Kong. Hasil penelitian tersebut menunjukkan bahwa luas bangunan dan usia properti menjadi faktor yang paling dominan dalam menentukan harga rumah. Selain itu, penggunaan *SHAP* juga terbukti mampu meningkatkan pemahaman dan kepercayaan pengguna terhadap model prediksi. Penelitian lain oleh Trindade Neves dan rekan-rekannya [11] mengombinasikan *XGBoost* dengan data terbuka serta teknik *XAI* untuk memprediksi harga properti di Lisbon, Portugal. Pendekatan tersebut berhasil meningkatkan performa model sekaligus memperjelas faktor-faktor yang memengaruhi prediksi harga. Sementara itu, Li [4] menemukan bahwa luas bangunan dan kondisi lingkungan sekitar menjadi variabel yang paling konsisten memengaruhi harga rumah pada berbagai dataset yang diuji.

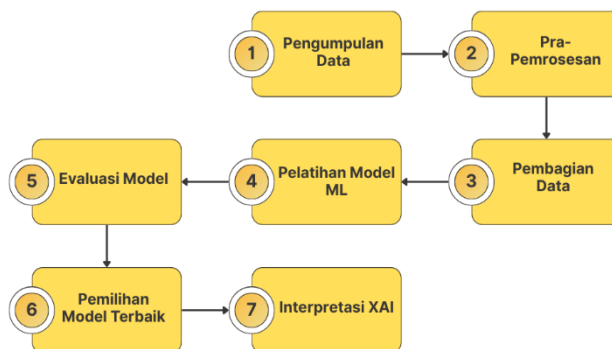
Dalam konteks Indonesia, penelitian mengenai prediksi harga rumah berbasis *machine learning* memang sudah mulai berkembang, namun sebagian besar masih berfokus pada peningkatan akurasi model. Emilita dan tim [3] menggunakan *Random Forest* untuk memprediksi harga properti berdasarkan karakteristik fisik dan lokasi, dan memperoleh nilai R^2 sebesar 0,8751. Meski demikian, penelitian tersebut belum membahas interpretasi model secara mendalam. Hal yang sama juga terlihat pada penelitian Putra dan Suhartana [12] yang menerapkan *Random Forest Regression* untuk prediksi harga rumah di Jabodetabek dengan hasil performa yang cukup tinggi, tetapi belum mengintegrasikan pendekatan *XAI* secara komprehensif.

Berdasarkan berbagai penelitian terdahulu, masih terlihat adanya kesenjangan penelitian, khususnya terkait integrasi antara akurasi prediksi dan interpretabilitas model dalam konteks data properti di Indonesia. Sebagian besar penelitian hanya berfokus pada peningkatan performa prediksi tanpa menjelaskan faktor-faktor yang memengaruhi hasil model secara transparan. Padahal, kebutuhan akan sistem yang tidak hanya akurat tetapi juga mudah dipahami semakin penting dalam pengambilan keputusan di sektor properti. Oleh karena itu, penelitian ini bertujuan membandingkan model prediksi harga rumah menggunakan empat algoritma *machine learning*, yaitu *Linear Regression*, *Random Forest*, *XGBoost*, dan *Gradient Boosting*, yang dipadukan dengan pendekatan *XAI* menggunakan *SHAP* dan *LIME*. Data yang digunakan berasal dari properti nyata di Indonesia dengan variabel utama berupa luas bangunan, luas tanah, jumlah kamar tidur, jumlah kamar mandi, dan kapasitas garasi. Melalui pendekatan ini, penelitian diharapkan mampu menghasilkan sistem prediksi harga rumah yang lebih akurat, transparan, dan dapat dipercaya oleh berbagai pihak yang terlibat dalam sektor properti di Indonesia.

2. METODOLOGI PENELITIAN

Metode penelitian yang digunakan adalah pendekatan kuantitatif berbasis eksperimen komputasi dengan mengimplementasikan empat algoritma *machine learning* serta teknik *Explainable Artificial Intelligence (XAI)*. Alur penelitian mengacu pada *framework Knowledge Discovery in Databases (KDD)* yang dimodifikasi, yang terdiri atas tahap

pengumpulan data, pra-pemrosesan, pelatihan model, evaluasi, dan interpretasi hasil menggunakan metode *XAI* untuk meningkatkan transparansi model prediksi.



Gambar 1. Alur Tahapan Penelitian

2.1 Pengumpulan Data

Dataset yang digunakan dalam penelitian ini adalah Daftar Harga Rumah yang bersumber dari platform Kaggle dan dapat diakses secara publik melalui tautan <https://www.kaggle.com/datasets/wisnuanggara/daftar-harga-rumah> [13]. Dataset ini berisi data harga properti residensial di kawasan Tebet, Jakarta Selatan, yang dikumpulkan dari berbagai sumber listingan properti. Data tersebut memuat 1.010 record dengan delapan atribut, yakni NO, NAMA RUMAH, LB, LT, KT, KM, GRS, HARGA. Atribut HARGA dijadikan sebagai variabel target, sedangkan LB, LT, KT, KM, dan GRS diposisikan sebagai variabel masukan (input features) pada model.

2.2 Pra-Pemrosesan

Sebelum data dimasukkan ke dalam model, dilakukan serangkaian proses pembersihan dan transformasi untuk memastikan kualitas data yang digunakan [14]. Atribut yang tidak diperlukan dihapus terlebih dahulu, kemudian nilai yang hilang (*missing values*) ditangani dengan cara mengganti nilai kosong menggunakan nilai median dari masing-masing atribut. Data yang mengandung nilai ekstrem (*outlier*) diidentifikasi dan dieliminasi, sehingga dari 1.010 data awal tersisa 918 data yang layak digunakan.

2.3 Pembagian Data

Data yang telah bersih kemudian dibagi menjadi dua kelompok, dengan komposisi 80% untuk keperluan pelatihan dan 20% sisanya untuk pengujian model [15]. Penetapan nilai random state = 42 dilakukan agar pembagian data bersifat deterministik dan hasil eksperimen dapat direproduksi kembali. Komposisi ini berlaku seragam untuk semua model yang dibangun.

2.4 Pelatihan Model ML

Penelitian ini membangun dan membandingkan empat model *machine learning*, yaitu *Linear Regression*, *Random Forest*, *XGBoost*, dan *Gradient Boosting*, yang seluruhnya diimplementasikan menggunakan library *scikit-learn* 1.2.0 dan *XGBoost* 1.7.0 di lingkungan Python 3.x. Keempat model dilatih pada data latih yang sama dengan konfigurasi hyperparameter masing-masing. *Linear Regression* diposisikan sebagai model pembanding dasar (*baseline*), sedangkan ketiga algoritma berbasis ensemble dirancang untuk menangkap hubungan yang lebih kompleks dan non-linier antara atribut fisik properti dengan harga jualnya.

2.5 Evaluasi Model

Setelah proses pelatihan selesai, performa setiap model diukur menggunakan empat metrik evaluasi regresi, yaitu *Mean Absolute Error (MAE)*, *Root Mean Squared Error (RMSE)*, koefisien determinasi (R^2), dan *Mean Absolute Percentage Error (MAPE)* [16]. Pengujian dilakukan secara eksklusif pada data uji yang tidak pernah digunakan selama pelatihan, sehingga hasil evaluasi mencerminkan kemampuan model yang sesungguhnya dalam menghadapi data baru.

2.6 Pemilihan Model Terbaik

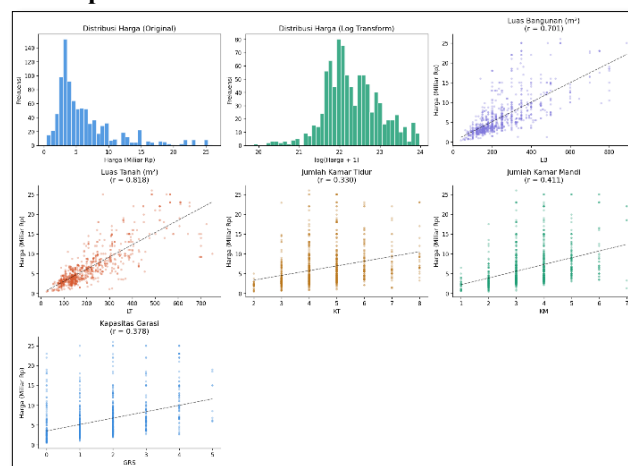
Model yang memberikan nilai R^2 tertinggi pada data uji ditetapkan sebagai model terbaik, sebab R^2 mencerminkan sejauh mana model mampu menerangkan keragaman nilai harga secara keseluruhan dan menjadi metrik utama yang lazim digunakan dalam berbagai penelitian perbandingan algoritma ensemble untuk prediksi harga properti [17]. Jika dua model memiliki selisih R^2 yang sangat kecil, nilai MAPE dijadikan pertimbangan tambahan dalam proses seleksi, sebagaimana yang diterapkan dalam studi perbandingan *Random Forest* dan *Gradient Boosting Machines* pada kasus serupa di wilayah Jakarta Selatan [18]. Model terpilih kemudian digunakan secara khusus sebagai objek analisis pada tahap interpretasi *XAI*, sehingga penjelasan yang dihasilkan benar-benar merepresentasikan perilaku model dengan performa terbaik [19].

2.7 Interpretasi *XAI*

Pada tahap ini, dua metode *XAI* diterapkan secara bersamaan terhadap model terbaik yang telah dipilih, yaitu *SHAP* (*SHapley Additive exPlanations*) dan *LIME* (*Local Interpretable Model-agnostic Explanations*). *SHAP* digunakan untuk menelaah kontribusi setiap fitur secara menyeluruh melalui *summary plot*, *bar chart*, dan *dependence plot*, sekaligus menjelaskan prediksi secara individual melalui *waterfall plot* pada tiga instance representatif, sebagaimana diterapkan secara efektif pada penelitian prediksi harga rumah berbasis *ensemble learning* untuk meningkatkan transparansi model [17]. *LIME* melengkapi analisis tersebut dengan membangun model pengganti yang lebih sederhana di sekitar setiap prediksi, sehingga mampu mengidentifikasi fitur-fitur kunci yang paling memengaruhi hasil prediksi pada level individual [20]. Kombinasi kedua metode *XAI* ini terbukti memberikan hasil interpretasi yang saling melengkapi, di mana *SHAP* digunakan untuk analisis global dan *LIME* untuk validasi lokal, sehingga menghasilkan gambaran interpretasi yang lebih utuh dan menyeluruh dibandingkan penggunaan salah satu metode secara tunggal [21].

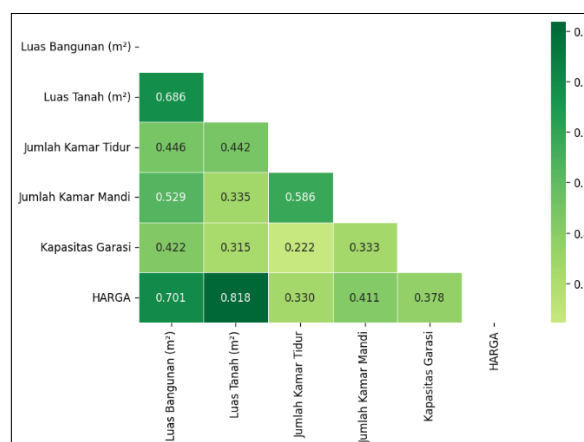
3. HASIL DAN PEMBAHASAN

3.1 Hasil Pra-Pemrosesan dan Eksplorasi Data



Gambar 2. Eksplorasi Dataset

Dari 1.010 data awal, sebanyak 92 data dihapus akibat teridentifikasi sebagai *outlier* ekstrem maupun memiliki nilai hilang, sehingga total data bersih yang digunakan dalam pemodelan berjumlah 918 data. Eksplorasi awal terhadap distribusi variabel target menunjukkan bahwa harga rumah memiliki pola sebaran yang sangat condong ke kanan (*right-skewed*), dengan sebagian besar properti terkonsentrasi di bawah Rp10 miliar dan sebagian kecil mencapai harga di atas Rp25 miliar. Kondisi ini diatasi dengan menerapkan transformasi logaritmik natural, yang berhasil mengubah distribusi harga menjadi lebih simetris dan mendekati kurva normal, sehingga lebih sesuai untuk proses pelatihan model regresi.

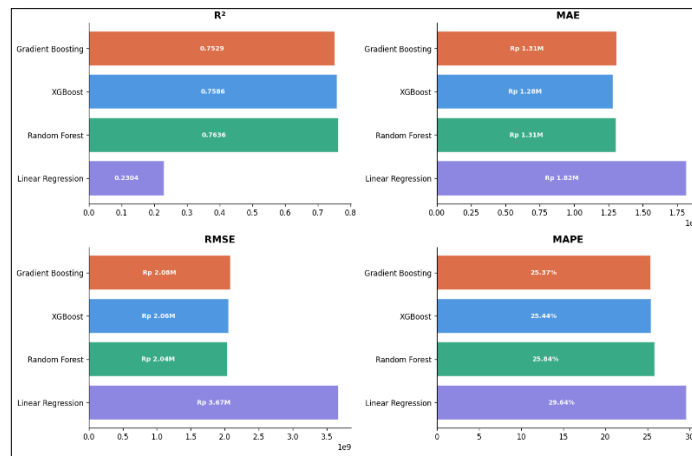


Gambar 3. Heatmap Korelasi Fitur vs Harga

Analisis korelasi Pearson antara fitur dan variabel target mengungkap bahwa Luas Tanah (LT) memiliki hubungan linear positif paling kuat dengan nilai $r = 0,818$, disusul Luas Bangunan (LB) dengan $r = 0,701$. Sementara itu, Jumlah Kamar Mandi (KM), Kapasitas Garasi (GRS), dan Jumlah Kamar Tidur (KT) masing-masing menunjukkan korelasi moderat

hingga lemah dengan nilai $r = 0,411$, $0,378$, dan $0,330$. Analisis *heatmap* antara sesama fitur juga menunjukkan adanya korelasi yang cukup signifikan antara LB dan LT ($r = 0,686$), serta antara KT dan KM ($r = 0,586$). Temuan ini mengindikasikan bahwa ukuran fisik properti, khususnya luas tanah dan luas bangunan, merupakan prediktor utama yang paling dominan dalam menentukan harga rumah pada dataset yang digunakan.

3.2 Perbandingan Performa Model *Machine Learning*



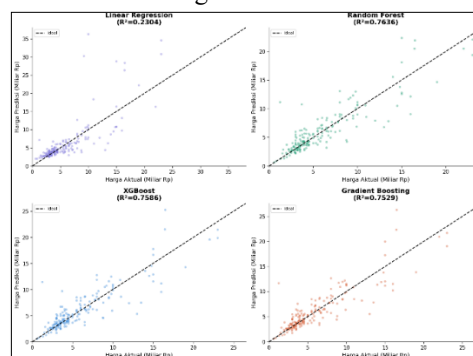
Gambar 4. Perbandingan Performa Model ML

Keempat model yang diuji menunjukkan hasil yang bervariasi secara signifikan. Tiga model berbasis *ensemble Random Forest*, *XGBoost*, dan *Gradient Boosting* secara konsisten mengungguli *Linear Regression* pada seluruh metrik evaluasi. Hasil lengkap perbandingan performa disajikan pada Tabel 1:

Tabel 1. Hasil Evaluasi Performa Keempat Model *Machine Learning*

Model	R^2	MAE (Rp)	RMSE (Rp)	MAPE (%)
<i>Random Forest</i>	0,7636	1,305 M	2,037 M	25,84
XGBoost	0,7586	1,285 M	2,058 M	25,44
<i>Gradient Boosting</i>	0,7529	1,311 M	2,083 M	25,37
<i>Linear Regression</i>	0,2304	1,819 M	3,675 M	29,64

Random Forest mencatat nilai R^2 tertinggi sebesar 0,7636, artinya model ini mampu menjelaskan sekitar 76,36% variansi harga rumah berdasarkan fitur yang tersedia. Nilai *RMSE*-nya juga paling rendah di antara seluruh model, yakni Rp2,037 miliar, yang menandakan konsistensi prediksi bahkan terhadap data dengan harga ekstrem. Meskipun *XGBoost* sedikit lebih unggul dari sisi *MAE* (Rp1,285 miliar) dan *Gradient Boosting* menghasilkan *MAPE* terendah (25,37%), selisih antar ketiga model *ensemble* ini tergolong kecil dan tidak substansial. Sebaliknya, *Linear Regression* menunjukkan performa yang jauh tertinggal dengan R^2 hanya 0,2304 dan *RMSE* mencapai Rp3,675 miliar hampir dua kali lipat dibandingkan model berbasis *ensemble*. Kondisi ini mempertegas bahwa pola harga properti dalam dataset bersifat non-linier dan tidak dapat direpresentasikan secara memadai oleh model regresi sederhana.



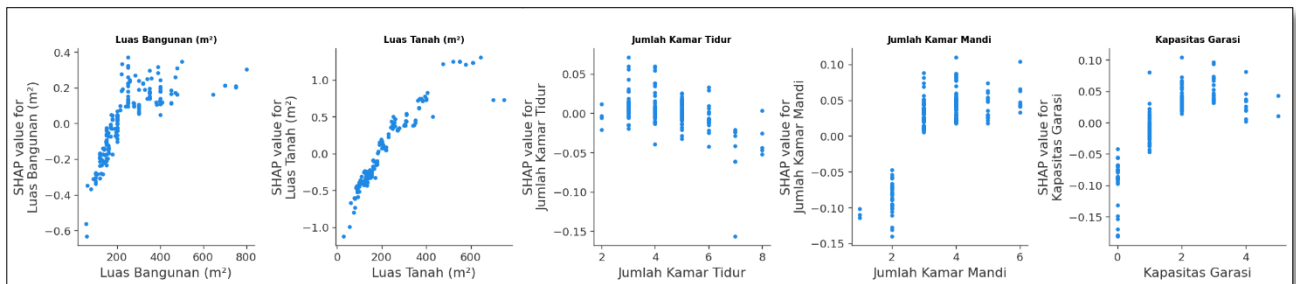
Gambar 5. Actual vs Predicted

Visualisasi *scatter plot actual vs. predicted* turut memperkuat kesimpulan tersebut. Pada *Random Forest*, titik-titik data tampak paling rapat mengikuti garis ideal ($y = x$) dibandingkan model lainnya. ketiga model *ensemble* menunjukkan peningkatan sebaran prediksi pada rentang harga premium di atas Rp15 miliar, yang mengindikasikan adanya keterbatasan model dalam menangkap karakteristik unik properti kelas atas yang kemungkinan dipengaruhi oleh faktor-

faktor di luar fitur yang tersedia. Berdasarkan pertimbangan menyeluruh terhadap keempat metrik, *Random Forest* ditetapkan sebagai model terbaik dan menjadi dasar analisis *XAI* pada tahap berikutnya.

3.3 Interpretasi Model Menggunakan SHAP

Analisis SHAP diterapkan pada model *Random Forest* untuk menghasilkan interpretasi secara global maupun lokal. *SHAP Feature Importance* menunjukkan bahwa Luas Tanah (LT) adalah fitur paling dominan dengan nilai *Mean SHAP* mendekati 0,40, jauh melampaui Luas Bangunan (LB) di posisi kedua dengan nilai sekitar 0,15. Fitur-fitur fasilitas seperti KM (0,05), GRS (0,04), dan KT (di bawah 0,02) memberikan kontribusi yang jauh lebih kecil. Pola ini mengonfirmasi bahwa model sangat bergantung pada dimensi fisik lahan dalam membentuk prediksi harga. *SHAP Summary Plot (Beeswarm)* memperlihatkan arah pengaruh yang konsisten: titik-titik dengan nilai LT dan LB tinggi (warna merah) terkonsentrasi di sisi kanan sumbu *SHAP*, yang berarti luas yang besar selalu mendorong prediksi harga ke atas. Sebaliknya, nilai LT dan LB rendah (warna biru) berada di sisi kiri dan cenderung menekan prediksi. Fenomena menarik ditemukan pada fitur KT, di mana sebagian titik merah justru berada di area negatif mengindikasikan bahwa jumlah kamar tidur yang berlebihan tanpa didukung luas area yang memadai tidak serta-merta menaikkan nilai properti.

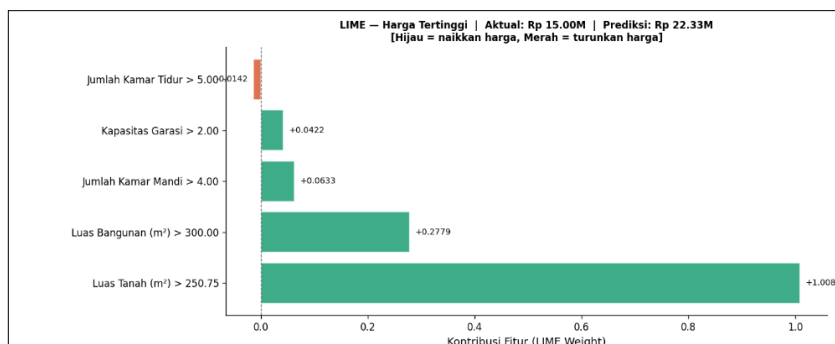


Gambar 6. SHAP Dependence Plot – Semua Fitur (Random Forest)

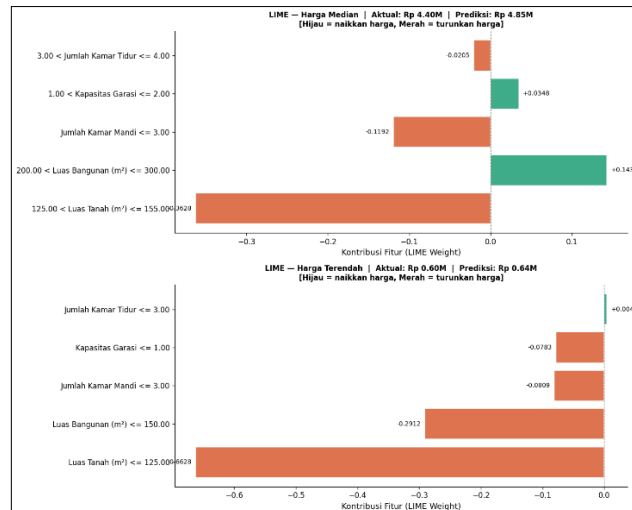
Analisis *SHAP Dependence Plot* memberikan gambaran yang lebih granular. Pada fitur LT, dampak positif mulai tampak signifikan setelah melewati angka 200 m², dengan rentang pengaruh SHAP terluas dari -1,0 hingga +1,0. Fitur LB menunjukkan pola serupa dengan efek percepatan yang mulai melandai (*plateau*) setelah 400 m². Pada KM ditemukan ambang batas yang jelas di angka 3 kamar mandi, di bawah nilai tersebut kontribusinya cenderung negatif, sedangkan pada GRS, ketiadaan garasi memberikan dampak negatif yang cukup besar dan baru terkompensasi secara positif saat kapasitas mencapai minimal 1 unit.

3.4 Interpretasi Model Menggunakan LIME

Analisis *LIME* pada tiga instance representatif menghasilkan penjelasan lokal yang konsisten dengan temuan *SHAP*. Pada rumah dengan harga aktual Rp15 miliar (prediksi Rp22,33 miliar), kondisi LT > 250,75 m² menjadi pendorong terkuat dengan bobot +1,0088, diperkuat oleh LB > 300 m² (+0,2779) dan KM > 4 (+0,0633). Menariknya, jumlah kamar tidur yang terlalu banyak (> 5) justru memberikan dampak negatif kecil sebesar -0,0142, selaras dengan temuan pada *SHAP Beeswarm* sebelumnya.



Gambar 7. LIME Local Explanations – Random Forest (1)



Gambar 8. LIME Local Explanations – Random Forest (2)

Pada rumah harga median dengan harga aktual Rp4,40 miliar (prediksi Rp4,85 miliar), faktor utama yang menekan harga adalah LT yang tergolong kecil antara 125–155 m² dengan dampak -0,3628, meskipun LB pada rentang 200–300 m² masih memberikan kontribusi positif sebesar +0,1431. Sementara pada rumah harga terendah (aktual Rp0,60 miliar, prediksi Rp0,64 miliar), hampir seluruh fitur berkontribusi negatif dengan LT ≤ 125 m² sebagai penyumbang terbesar (-0,6628), diikuti LB ≤ 150 m² (-0,2912) dan minimnya fasilitas pendukung.

Secara keseluruhan, terdapat konsistensi yang kuat antara hasil *SHAP* dan *LIME* dalam menempatkan LT dan LB sebagai dua faktor penentu utama di seluruh rentang harga. Konsistensi antara kedua metode *XAI* yang memiliki pendekatan berbeda ini memperkuat validitas interpretasi model. *SHAP* memberikan gambaran menyeluruh tentang kepentingan fitur secara global, sedangkan *LIME* melengkapinya dengan penjelasan yang lebih kontekstual pada level prediksi individual, sehingga keduanya bersama-sama menghasilkan pemahaman yang lebih utuh terhadap mekanisme kerja model dalam memprediksi harga rumah.

4. KESIMPULAN

Penelitian ini berhasil membangun model prediksi harga rumah berbasis *machine learning* yang dipadukan dengan analisis *Explainable Artificial Intelligence (XAI)* menggunakan dataset properti kawasan Tebet, Jakarta Selatan. Dari empat algoritma yang diuji, *Random Forest* terbukti menjadi model terbaik dengan nilai $R^2 = 0,7636$, MAE = Rp1,305 miliar, RMSE = Rp2,037 miliar, dan MAPE = 25,84%, mengungguli *XGBoost*, *Gradient Boosting*, maupun *Linear Regression* yang mencatat R^2 terendah sebesar 0,2304 membuktikan bahwa hubungan antara fitur fisik properti dan harga bersifat non-linier sehingga tidak dapat ditangkap secara memadai oleh model regresi sederhana. Analisis *SHAP* mengungkap bahwa Luas Tanah merupakan faktor penentu harga paling dominan dengan nilai Mean *SHAP* mendekati 0,40, disusul Luas Bangunan, sementara fitur fasilitas seperti jumlah kamar dan kapasitas garasi berperan sebagai pendukung dengan pengaruh yang jauh lebih kecil. Hasil analisis *LIME* pada tiga instance representatif menunjukkan konsistensi yang tinggi dengan temuan *SHAP*, di mana Luas Tanah secara konsisten menjadi faktor pendorong atau penekan harga terbesar pada setiap prediksi individual yang dianalisis. Konsistensi antara kedua metode *XAI* ini memperkuat validitas interpretasi model secara keseluruhan dan membuktikan bahwa pendekatan kombinasi *machine learning* dan *XAI* tidak hanya menghasilkan prediksi yang lebih akurat dibandingkan metode konvensional, tetapi juga mampu memberikan penjelasan yang transparan dan dapat dipahami oleh berbagai pemangku kepentingan di sektor properti. Untuk penelitian selanjutnya, penambahan fitur lokasi seperti koordinat geografis dan jarak ke fasilitas publik, serta perluasan cakupan dataset, berpotensi meningkatkan performa model secara signifikan.

UCAPAN TERIMAKASIH

Penelitian ini terlaksana atas dukungan pendanaan dari Lembaga Penelitian dan Pengabdian kepada Masyarakat (LPPM) Universitas Negeri Medan (UNIMED) melalui Program Hibah Penelitian Dosen Pemula Internal UNIMED Tahun Anggaran 2026. Penulis mengucapkan terima kasih yang sebesar-besarnya atas kepercayaan, dukungan, dan fasilitas yang diberikan sehingga penelitian ini dapat diselesaikan dengan baik.

REFERENCES

- [1] N. A. Cahyani Putri and D. B. Arianto, "Komparasi Penggunaan Information Gain Pada Machine Learning untuk Memprediksi Harga Rumah di Jabodetabek," *J. Sains dan Teknol.*, vol. 5, no. 3, pp. 756–762, 2024, doi: 10.55338/saintek.v5i3.2052.
- [2] E. Fitri, "Analisis Perbandingan Metode Regresi Linier, Random Forest Regression dan Gradient Boosted Trees Regression Method untuk Prediksi Harga Rumah," *J. Appl. Comput. Sci. Technol.*, vol. 4, no. 1, pp. 58–64, 2023, doi:

- 10.52158/jacost.v4i1.491.
- [3] S. Emilita, P. S. Rahayu, S. R. Larasati, G. Taufiq, and J. T. Kumalasari, “Penerapan Algoritma Random Forest Dalam Memprediksi Harga Properti di Jakarta Selatan Berdasarkan Karakteristik Fisik dan Lokasi,” *J. Minfo Polgan*, vol. 14, no. 2, pp. 3415–3421, 2025, doi: 10.33395/jmp.v14i2.15735.
 - [4] H. Li, “House Price Prediction and Analysis Based on Random Forest and XGBoost Models,” *Highlights Business, Econ. Manag.*, vol. 21, pp. 934–938, 2023, doi: 10.54097/hbem.v2i1.14837.
 - [5] A. Fauzi, N. Maulidah, R. Supriyadi, H. Nalatissifa, and S. Diantika, “Prediksi Harga Properti Di Indonesia Menggunakan Algoritma Random Forest,” *RIGGS J. Artif. Intell. Digit. Bus.*, vol. 2, no. 2, 2024, doi: 10.61233/riggs.v2i2.367.
 - [6] A. N. Rais, Warjiyono, I. Alfarobi, S. W. Hadi, and W. Kurniawan, “Analisa Prediksi Harga Jual Rumah Menggunakan Algoritma Random Forest Machine Learning,” *JURSISTEKNI (Jurnal Sist. Inf. dan Teknol. Informasi)*, vol. 6, no. 2, pp. 416–423, 2024, doi: 10.52005/jursistekni.v6i2.323.
 - [7] B. Wulan Sari and D. Prabowo, “Analisis Perbandingan Prediksi Harga Rumah Dengan Random Forest, Gradient Boosting, dan XGBoost,” *J. Learn. Technol. Innov.*, vol. 4, no. 1, pp. 42–51, 2025, doi: 10.57255/intellect.v4i1.1385.
 - [8] M. Sajid, K. Hussain, S. Hussain, and A. Hussain, “Interpreting Black-Box Models: A Review on Explainable Artificial Intelligence,” *Cognit. Comput.*, vol. 16, no. 1, pp. 45–74, 2024, doi: 10.1007/s12559-023-10179-8.
 - [9] I. Lenaers and L. De Moor, “Exploring XAI Techniques for Enhancing Model Transparency and Interpretability in Real Estate Rent Prediction: A Comparative Study,” *Financ. Res. Lett.*, vol. 58, p. 104306, 2023, doi: 10.1016/j.frl.2023.104306.
 - [10] W. K. O. Ho, “eXplainable Machine Learning for Real Estate: XGBoost and Shapley Values in Price Prediction,” *Civ. Eng. J.*, vol. 11, no. 5, pp. 2116–2133, 2025, doi: 10.28991/CEJ-2025-011-05-022.
 - [11] F. Trindade Neves, M. Aparicio, and M. de Castro Neto, “The Impacts of Open Data and eXplainable AI on Real Estate Price Predictions in Smart Cities,” *Appl. Sci.*, vol. 14, no. 5, p. 2209, 2024, doi: 10.3390/app14052209.
 - [12] I. M. G. A. B. Putra and I. K. G. Suhartana, “Implementasi Algoritma Random Forest Regression dalam Sistem Prediksi Harga Rumah di Jabodetabek,” *J. Nas. Teknol. Inf. dan Apl.*, vol. 4, no. 1, pp. 27–38, 2025, doi: 10.24843/JNATIA.2025.v04.i01.p04.
 - [13] W. Anggara, “Daftar Harga Rumah.” Kaggle, 2020. [Online]. Available: <https://www.kaggle.com/datasets/wisnuanggara/daftar-harga-rumah>
 - [14] D. Praseptian M, Sinawati, and K. Harianto, “Perbandingan Penanganan Missing Value pada Data Numerik Survei Kepuasan Pengguna Lulusan,” *Decod. J. Pendidik. Teknol. Inf.*, vol. 5, no. 2, pp. 789–798, 2025, [Online]. Available: <https://journal.umkendari.ac.id/decode/article/view/1262>
 - [15] D. Benaya, “Implementasi Random Forest dalam Klasifikasi Kanker Paru-Paru,” *JOINTER J. Informatics Eng.*, vol. 5, no. 01, pp. 27–31, 2024, doi: 10.53682/jointer.v5i01.331.
 - [16] M. B. S. Qolbi, T. N. Puteh, R. Rivandi, and C. Rozikin, “Prediksi Harga Rumah di Jakarta Pusat Menggunakan Algoritma Machine Learning,” *J. Ilmu Komput. dan Bisnis*, vol. 16, no. 1, pp. 16–24, 2025, doi: 10.47927/jikb.v16i1.840.
 - [17] Dhiwa Aqsha, “Perbandingan Kinerja Algoritma Extreme Gradient Boosting Dan Random Forest Untuk Prediksi Harga Rumah Di Jabodetabek,” *J. Ilmu Komput. dan Sist. Inf.*, vol. 13, no. 1, pp. 1–7, 2025, doi: 10.24912/jiksi.v13i1.32863.
 - [18] F. A. Rangkuti, S. Sundari, T. Informatika, and U. H. Medan, “Implementasi Gradient Boosting Machines Untuk Prediksi Harga Implementation of Gradient Boosting Machines for House Price Prediction in South Jakarta,” *J. Kecerdasan Buatan dan Teknol. Inf.*, vol. 4, no. 2, pp. 164–172, 2025.
 - [19] A. P. Wardani, H. A. Irawan, M. P. Syah, M. A. Akmal, N. U. Nariswari, and K. M. Hindrayani, “Analisis dan Prediksi Harga Properti Rumah di Kota Surabaya dengan Algoritma Random Forest,” *Pros. Semin. Nas. Sains Data*, vol. 4, no. 1, pp. 885–894, 2024, doi: 10.33005/senada.v4i1.375.
 - [20] Mubaqarah, A. N. Puteri, and A. Sumardin, “Comparison of Random Forest and XGBoost for Diabetes Classification with SHAP and LIME Interpretation,” *JTERA (Jurnal Teknol. Rekayasa)*, vol. 9, no. May, pp. 121–130, 2024, doi: 10.31544/jtera.v9.i1.2024.121-130.
 - [21] M. N. Hibatulloh, G. D. Prakoso, A. D. Putri Yunus, and T. D. Putra, “Prediksi Harga Rumah di Bandung 2024 Menggunakan Ensemble Learning: Analisis Komparatif dan Interpretabilitas,” *J. Inform. J. Pengemb. IT*, vol. 10, no. 2, pp. 484–493, 2025, doi: 10.30591/jpit.v10i2.8200.