Jurnal Informatika: Jurnal pengembangan IT

Vol. 10, No. 2, 2025
ISSN : 2477-5126, DOI:10.30591/jpit.v9ix.xxx

Prediksi Harga Rumah di Bandung 2024 Menggunakan Ensemble Learning: Analisis Komparatif dan Interpretabilitas

Muh Naufal Hibatulloh[1], Gilang Danu Prakoso,[2] Adiestiana Dwi Putri Yunus[3], Tommy Dwi Putra[4]
[1,2,3,4]STMIK Amikom Surakarta, Jl. Veteran, Dusun I, Singopuran, Kec. Kartasura, Kabupaten Sukoharjo, Jawa Tengah 57163

Info Artikel

Riwayat Artikel:

Received 2025-01-02
Revised 2025-04-29
Accepted 2025-05-15

Corresponding Author:

Muh Naufal Hibatulloh

Email:

muhnaufalhibatulloh@gmail.com

This is an open access article under the CC BY 4.0 license.

Abstract *– House price prediction plays a crucial role in investment decision-making and financial planning, particularly in developing cities like Bandung with its complex property market dynamics. This study aims to evaluate and compare the performance of various ensemble learning techniques in predicting house prices in Bandung for the year 2024, with a specific focus on model interpretability analysis. The data was collected through web scraping from www.rumah123.com in March 2024, covering attributes such as location, number of rooms, land area, and building area. The evaluated ensemble techniques include Random Forest, Gradient Boosting Machines, Xtreme Gradient Boosting, Linear Regression, and Stacking Ensemble. Model performance was assessed using MAE, RMSE, and R-squared metrics, while interpretability analysis was conducted using SHAP values. The Model Stacking Ensemble shows the most optimal results with R^2 0.9076, RMSE 0.311, and MAE 0.216 in experiments involving location features. Features such as land size, building size, and location have proven to have the greatest impact in predicting prices based on SHAP analysis. This model has been successfully integrated into a Flask website for interactive price predictions.*

Keywords: *Ensemble Learning; House Price Prediction; SHAP; Model Interpretability; Bandung.*

Abstrak *– Prediksi harga rumah memainkan peran krusial dalam pengambilan keputusan investasi dan perencanaan keuangan, terutama di kota berkembang seperti Bandung dengan dinamika pasar properti yang kompleks. Penelitian ini bertujuan untuk mengevaluasi dan membandingkan kinerja berbagai teknik ensemble learning dalam memprediksi harga rumah di Bandung untuk tahun 2024, dengan fokus khusus pada analisis interpretabilitas model. Data diperoleh melalui web scraping dari www.rumah123.com pada Maret 2024, mencakup atribut seperti lokasi, jumlah kamar, luas tanah, dan luas bangunan. Teknik ensemble yang dievaluasi meliputi Random Forest, Gradient Boosting Machines, Xtreme Gradient Boosting, Linear Regression, dan Stacking Ensemble. Kinerja model dinilai menggunakan metrik MAE, RMSE, dan R-squared, sementara analisis interpretabilitas dilakukan menggunakan teknik SHAP values. Model Stacking Ensemble menunjukkan hasil paling optimal dengan R^2 0,9076, RMSE 0,311, dan MAE 0,216 dalam percobaan yang melibatkan fitur lokasi. Fitur-fitur seperti ukuran tanah, ukuran bangunan, dan lokasi terbukti memiliki dampak terbesar dalam memprediksi harga berdasarkan analisis SHAP. Model ini telah berhasil diintegrasikan ke dalam situs web Flask untuk melakukan prediksi harga secara interaktif.*

Kata Kunci*: Ensemble Learning, Prediksi Harga Rumah, SHAP, Interpretabilitas Model, Bandung*

I. PENDAHULUAN

Prediksi harga rumah telah menjadi fokus utama dalam penelitian real estate, mengingat perannya yang krusial dalam pengambilan keputusan investasi dan perencanaan keuangan [1], [2]. Di kota-kota berkembang seperti Bandung, Indonesia, dinamika pasar properti yang kompleks menimbulkan tantangan unik dalam memprediksi harga rumah. Dengan pertumbuhan ekonomi yang pesat dan urbanisasi yang terus meningkat [2], sehingga tentunya menjadikan pasar properti Bandung mengalami fluktuasi signifikan dalam beberapa tahun terakhir.

Berbagai metode telah digunakan untuk memprediksi harga rumah [3], [4], mulai dari teknik statistik tradisional hingga algoritma *machine learning* yang lebih canggih. Metode seperti regresi linier [3], [5], [6], decision tree [7], dan random forest [7] telah banyak diaplikasikan dalam konteks ini. Namun, masing-masing metode ini memiliki keterbatasan dalam menangkap kompleksistas faktor-faktor yang mempengaruhi harga rumah, terutama dalam memprediksi harga untuk periode tertentu.

Dalam beberapa tahun terakhir, teknik ensemble learning [8] telah menunjukkan potensi dalam meningkatkan akurasi prediksi di berbagai domain[9], termasuk prediksi harga properti. Ensemble learning, yang menggabungkan prediksi dari beberapa model untuk menghasilkan hasil yang lebih akurat dan robust, menawarkan pendekatan yang menjanjikan untuk mengatasi keterbatasan model tunggal [10]. Namun, penerapan dan efektivitas teknik ini dalam konteks pasar properti di kota berkembang seperti Bandung masih perlu dieksplorasi lebih lanjut.

Selain akurasi, interpretabilitas model menjadi aspek penting dalam konteks prediksi harga rumah [1], [2]. Model yang dapat menjelaskan kontribusi relatif dari faktor-faktor seperti lokasi, jumlah kamar tidur, atau luas tanah tidak hanya meningkatkan kepercayaan pengguna tetapi juga memberikan wawasan berharga untuk

pengambilan keputusan strategis [2]. Namun, sering kali terdapat *trade-off* antara kompleksitas model *ensemble* dan interpretabilitasnya yang perlu dieksplorasi lebih lanjut.

Penelitian ini bertujuan untuk mengevaluasi dan membandingkan berbagai teknik ensemble learning dalam memprediksi harga rumah di Bandung untuk tahun 2024, dengan fokus khusus pada analisis interpretabilitas model. Secara spesifik, penelitian ini bertujuan untuk: [1] Mengembangkan model prediksi harga rumah di Bandung menggunakan berbagai teknik *ensemble learning*; [2] Melakukan analisis komparatif [11] terhadap kinerja berbagai teknik *ensemble*; [3] Mengevaluasi interpretabilitas model-model yang dihasilkan [1], [2]; dan [4] Memberikan rekomendasi untuk pengembangan model prediksi harga rumah yang akurat dan interpretable di pasar properti berkembang.

Metodologi penelitian ini akan menggunakan Dataset Harga Rumah Di Bandung tahun 2024 yang diperoleh dari kaggle.com. Dataset ini merupakan hasil web scraping dari www.rumah123.com pada bulan Maret 2024 dan berfokus pada properti residensial di Bandung, Jawa Barat, Indonesia. Dataset yang telah dibersihkan mencakup atribut-atribut seperti nama rumah, lokasi, jumlah kamar tidur, jumlah kamar mandi, jumlah carport, harga, luas tanah, dan luas bangunan.

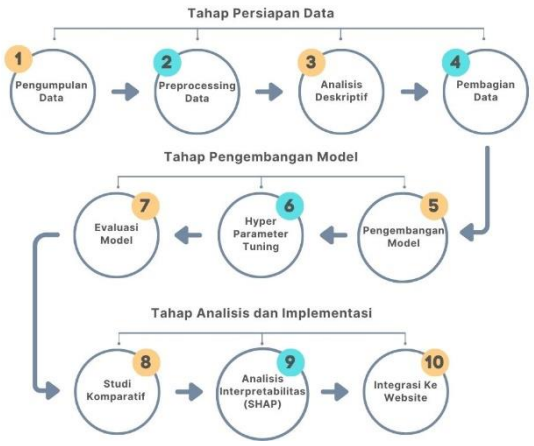
Teknik *ensemble* yang akan dievaluasi meliputi *Random Forest* [7], [12], *Gradient Boosting Machines* [13],*Xtreme Gradient Boosting*, *Linear Regression* dan *Stacking Ensemble*. Kinerja model akan dinilai menggunakan metrik seperti *Mean Absolute Error* (MAE), *Root Mean Squared Error* (RMSE) dan *R-squared*. Untuk analisis interpretabilitas, kami akan menggunakan teknik SHAP (*SHapley Additive exPlanations*) values [14], [15] untuk mengidentifikasi kontribusi relatif dari berbagai fitur terhadap prediksi harga.

Penelitian ini memberikan sumbangan akademis dengan meningkatkan wawasan tentang keefektifan metode *ensemble learning* di pasar real estate kota yang sedang tumbuh, serta mengkaji perimbangan antara ketepatan model dan kemudahan pemahaman. Dari segi praktis, hasil ini bisa mendukung para pengambil keputusan dalam urusan pasar properti di Bandung dan menyediakan suatu kerangka metodologi yang dapat digunakan untuk meramalkan harga properti di kota-kota berkembang yang lain.

II. METODE

Pada penelitian ini menggunakan pendekatan kuantitatif dengan desain studi komparatif yaitu untuk memprediksi harga rumah di Kota Bandung tahun 2024 menggunakan teknik *Ensemble Learning*. Adapun fokus penelitian kali ini yaitu membandingkan kinerja berbagai model prediksi dan menganalisis interpretabilitasnya. Untuk menjamin kebenaran dan validitas hasil prediksi harga rumah di Bandung pada tahun 2024, penelitian ini akan melakukan beberapa tahap utama. Tahap-tahap ini termasuk pengumpulan data, *preprocessing data*, pembagian data, pengembangan model, studi komparatif, dan analisis interpretabilitas.

Berikut gambar alur metode penelitian kali ini.



Gambar 1. Diagram Alur Metode Penelitian

A. Pengumpulan Data

Data yang digunakan dalam penelitian ini berasal dari dataset Harga Rumah di Bandung tahun 2024, yang dapat diakses melalui Kaggle.com. Dataset ini berasal dari web scraping yang dilakukan pada bulan Maret 2024 dari situs www.rumah123.com. Data yang dikumpulkan berfokus pada properti rumah di Bandung, Jawa Barat, Indonesia. Data tersebut mencakup informasi penting seperti lokasi, nama rumah, jumlah kamar tidur, jumlah kamar mandi, harga, luas tanah, dan luas bangunan.

B. Preprocessing Data

Sebelum data digunakan dalam pemodelan, langkah preprocessing dilakukan untuk membersihkan dan mempersiapkannya. Pada proses ini, yaitu melihat apakah ada data yang tidak ada nilainya, apakah ada data duplikat, apakah ada outliers pada dataset yang harus diatasi, normalisasi data yaitu pada data label price atau harga rumah karena rentang nilainya sangat jauh dengan variable lainnya. Data yang tidak lengkap atau tidak valid dan data duplikat akan dihapus atau diimputasi dengan menggunakan teknik yang sesuai.Data *Outliers* pada dataset juga diatasi dengan Teknik yang sesuai. *Preprocessing* juga melibatkan pengkodean variabel kategorikal menjadi format numerik yaitu pada variabel lokasi sehingga algoritma pembelajaran mesin dapat menggunakannya.

C. Analisis Deskriptif dan Pembagian Data

Setelah data diproses, langkah berikutnya adalah mendeskripsikan dan membaginya menjadi dua kelompok utama: data pelatihan (*training data*) dan data pengujian (*testing data*). Pembagian ini biasanya dilakukan dengan rasio 80:20 atau 70:30, tergantung pada ukuran dan fitur dataset. Data pelatihan nanti digunakan untuk membangun model prediksi, sedangkan data pengujian akan digunakan untuk mengevaluasi kinerja model yang telah dibangun.

D. Tahap Pengembangan Model

Penelitian ini akan menyelidiki tiga pendekatan pembelajaran yang populer: *Random Forest Regressor, Gradient Boosting Regressor, Xtreme Gradient Boosting Regressor (XGBoost), Linear Regression* dan *Stacking.*

1) Random Forest Regressor

Random Forest adalah metode *ensemble* yang menggunakan *multiple decision trees* untuk menghasilkan prediksi. Setiap tree atau pohon dilatih pada subset data yang berbeda menggunakan teknik bagging (*bootstrap aggregating*). Terlihat pada rumus nomor 1 yaitu *Random Forest Regressor* dimana ŷ adalah prediksi akhir, n adalah jumlah trees, dan hi(x) adalah prediksi dari tree ke-i.

$$\hat{y} = 1/n \sum(i=1 \text{ to } n)\ hi(x) \quad (1)$$

2) Gradient boosting Regressor

Gradient Boosting adalah Teknik yang membangun model secara bertahap, dimana setiap model baru berusaha memperbaiki keslahan dari model sebelumnya. Terlihat pada rumus nomor 2 yaitu *Gradient Boosting Regressor* dimana ŷi adalah prediksi untuk observasi i, K adalah jumlah pohon, dan fk merepresentasikan pohon ke-k.

$$F(x) = \sum(i=1 \text{ to } m)\ \gamma i\ hi(x) \quad (2)$$

3) XGBoost Regressor

XGBoost (Extreme Gradient Boosting) adalah implementasi yang dioptimalkan dari *Gradient Boosting. XGBoost* menggunakan regularisasi untuk mengurangi overfitting dan menangani data yang hilang secara native. Terlihat pada rumus nomor 3 *XGBoost Regressor* dimana ŷi adalah prediksi untuk observasi i, K adalah jumlah pohon, dan fk merepresentasikan pohon ke-k.

$$\hat{y}i = \sum(k=1 \text{ to } K)\ fk(xi) \quad (3)$$

4) *Linear Regression*

Linear Regression adalah model dasar yang memodelkan hubungan linear antara variabel input dan output. Terlihat pada rumus nomor 4 *Linear Regression* dimana y adalah variabel target, xi adalah variabel prediktor, βi adalah koefisien, dan ε adalah error term.

$$y = \beta 0 + \beta 1x1 + \beta 2x2 + \ldots + \beta ixi \quad (4)$$

5) *Stacking*

Stacking adalah metode ensemble yang menggabungkan prediksi dari beberapa model dasar menggunakan model meta. Terlihat pada rumus nomor 5 *Stacking* dimana ŷ adalah prediksi akhir, fmeta adalah model meta, dan fi(x) adalah prediksi dari model dasar ke-i.

$$\hat{y} = \text{fmeta}(\text{f1}(x), \text{f2}(x), ..., \text{fi}(x)) \quad (5)$$

Untuk membangun model prediksi harga rumah, masing-masing model ini akan diterapkan pada data pelatihan. Dengan tujuan untuk meningkatkan akurasi dan mengurangi overfitting.

E. Studi Komparatif

Secara menyeluruh, kinerja model yang dibangun menggunakan berbagai pendekatan pembelajaran kelompok akan dibandingkan. Untuk melakukan evaluasi ini, metrik kinerja seperti rata-rata kesalahan absolut (MAE), *Root Mean Squared Error* (RMSE) dan *R-squared* (R2) akan digunakan. MAE mengukur rata-rata kesalahan absolut antara nilai aktual dan prediksi, sedangkan R2 mengukur seberapa baik variabel independen menjelaskan variabilitas dalam variabel dependen. Tujuan dari studi komparatif ini adalah untuk menemukan metode kelompok yang paling tepat untuk memprediksi harga rumah di Bandung.

F. Analisis Interpretabilitas

Untuk memahami kontribusi relatif dari berbagai fitur terhadap prediksi harga rumah, analisis interpretabilitas model dilakukan. Nilai SHAP (*SHapley Additive exPlanations*) akan digunakan untuk menentukan dan menjelaskan dampak setiap fitur pada output model. Nilai SHAP memberikan nilai penting yang dapat diinterpretasikan baik di tingkat lokal maupun internasional, sehingga pengguna dapat memahami alasan di balik prediksi tertentu. Analisis ini akan meningkatkan kepercayaan pengguna terhadap model dan memberikan informasi penting untuk pengambilan keputusan strategis.

G. Integrasi ke Website

Integrasi ke website akan menggunakan Flask sebagai server backend. Proses ini meliputi:

- Frontend dengan membuat halaman web sederhana dengan HTML dan CSS, termasuk form input untuk fitur rumah
- Backend dengan mengembangkan app.py sebagai server Flask, yang akan menangani permintaan dari frontend, memuat model prediksi, memproses input dan mengembalikan hasil prediksi
- Integrasi dengan menghubungkan frontend dan backend menggunakan metode POST HTML.
- Pengujian dengan menjalankan dan menguji aplikasi secara lokal untuk memastikan fungsionalitas.

Pendekatan ini memungkinkan pengguna memasukkan data rumah melalui antarmuka web dan menerima prediksi harga secara langsung.

III. HASIL DAN PEMBAHASAN

A. Analisis Data dan Preprocessing

1) *Karakteristik Dataset:* Analisis awal dilakukan dengan melihat statistik deskriptif dari dataset Harga Rumah Bandung. Hasil analisis statistik deskriptif disajikan sebagai berikut:

	bedroom_count	bathroom_count	carport_count	price	land_area	building_area (m2)
count	7609.0000	7609.0000	7609.0000	7609.0000	7609.0000	7609.0000
mean	4.4769	3.1965	1.3747	4981946576.4227	274.5539	642.4599
std	4.7125	4.1048	1.9281	14762567073.4327	440.5538	34458.9781
min	0.0000	0.0000	0.0000	12000000.0000	6.0000	-250.0000
25%	3.0000	2.0000	0.0000	1150000000.0000	102.0000	100.0000
50%	4.0000	3.0000	1.0000	2250000000.0000	165.0000	180.0000
75%	5.0000	3.0000	2.0000	4750000000.0000	313.0000	300.0000
max	99.0000	99.0000	40.0000	735000000000.0000	11850.0000	3006000.0000

Gambar 2. Statistik Deskriptif Awal

Dari hasil analisis, terlihat bahwa beberapa data memiliki nilai yang kurang masuk akal. Contohnya, terdapat nilai negatif pada building_area, serta jumlah bathroom_count dan bedroom_count yang mencapai 99 kamar. Hal ini mengindikasikan perlunya dilakukan preprocessing data lebih lanjut.

2) *Preprocessing Data:* Tahap *preprocessing data* meliputi penanganan missing values, data duplikat, dan *outliers*. Setelah dilakukan analisis, tidak ditemukan missing values dalam dataset. Namun, terdapat 617 data duplikat, beberapa *outliers* dengan nilai ekstrem, serta nilai negatif yang seharusnya tidak ada. Langkah-langkah *preprocessing* yang dilakukan yaitu penghapusan data duplikat lalu penghapusan outliers menggunakan teknik *winsorizing*.

Winsorizing adalah teknik yang mengurangi dampak dari outliers dengan memangkas data di kedua ujung distribusi dan menggantikannya dengan nilai batas tertentu, dalam penelitian ini menggunakan persentil ke-1 dan ke-99.

	bedroom_count	bathroom_count	carport_count	price	land_area	building_area (m2)
count	6627.0000	6627.0000	6627.0000	6627.0000	6627.0000	6627.0000
mean	4.1266	2.8895	1.2998	4084494492.2288	239.5843	228.2947
std	2.8392	2.1058	1.4617	5455642970.5339	207.5445	176.7799
min	0.0000	0.0000	0.0000	240000000.0000	48.0000	36.0000
25%	3.0000	2.0000	0.0000	1170000000.0000	103.0000	100.0000
50%	4.0000	3.0000	1.0000	2200000000.0000	163.0000	180.0000
75%	5.0000	3.0000	2.0000	4500000000.0000	300.0000	300.0000
max	24.0000	20.0000	10.0000	40000000000.0000	1500.0000	1100.0000

Gambar 3. Statistik Deskriptif Setelah Winsorizing

Setelah itu, dilakukan analisis distribusi menggunakan *skewness* pada data. *Skewness* adalah ukuran statistik yang menggambarkan asimetri distribusi data. Dalam konteks analisis data, skewness membantu kita memahami sejauh mana data menyimpang dari distribusi normal. Hasil analisis menunjukkan bahwa sebagian besar fitur memiliki *skewness* ke kanan, namun masih dalam batas yang wajar.

```
Skewness:
 bedroom_count                      3.2210
bathroom_count                      3.6599
carport_count                       2.4780
price                               3.2750
land_area                           2.1830
building_area (m2)                  1.7084
building_to_land_ratio              1.1477
location_Antapani, Bandung          5.0488
location_Arcamanik, Bandung         3.9597
location_Astanaanyar, Bandung       4.3840
location_Babakanciparay, Bandung    4.5154
location_Bandung Kidul, Bandung     5.1606
location_Bandung Kulon, Bandung     5.2522
location_Bandung Wetan, Bandung     4.9203
location_Batununggal, Bandung       4.2697
location_Bojongloa Kidul, Bandung   4.8105
location_Buah Batu, Bandung         4.3097
location_Cibeunying Kidul, Bandung  6.5380
location_Cibiru, Bandung            4.7068
location_Cicendo, Bandung           4.2384
location_Cidadap, Bandung           4.6768
location_Coblong, Bandung           6.3750
location_Gede Bage, Bandung         5.8544
location_Kiaracondong, Bandung      5.8189
location_Lengkong, Bandung          4.2936
location_Mandalajati, Bandung       4.1112
location_Panyileukan, Bandung       5.5704
location_Rancasari, Bandung         5.1863
location_Regol, Bandung             4.9893
location_Sukajadi, Bandung          6.0997
location_Sukasari, Bandung          5.5704
location_Sumurbandung, Bandung      5.4051
location_Ujungberung, Bandung       5.3203
```

Gambar 4. Analisis Skewness

Selanjutnya, dilakukan analisis korelasi setiap atribut terhadap variabel harga (*price*) menggunakan korelasi *Pearson*. Hasil analisis menunjukkan bahwa land_area memiliki korelasi tertinggi terhadap variabel *price*, diikuti oleh *building_area* dan beberapa fitur lokasi.

```
land_area                              0.7763
building_area (m2)                     0.6725
location_Bandung Wetan, Bandung        0.3611
carport_count                          0.3541
location_Sumurbandung, Bandung         0.2820
bathroom_count                         0.2600
bedroom_count                          0.2431
location_Coblong, Bandung              0.1720
location_Cidadap, Bandung              0.1080
location_Cicendo, Bandung              0.0961
location_Sukajadi, Bandung             0.0761
location_Lengkong, Bandung             0.0566
location_Sukasari, Bandung             0.0423
location_Batununggal, Bandung          0.0175
location_Bojongloa Kidul, Bandung     -0.0012
location_Bandung Kidul, Bandung       -0.0236
location_Regol, Bandung               -0.0303
location_Babakanciparay, Bandung      -0.0315
location_Cibeunying Kidul, Bandung    -0.0395
location_Kiaracondong, Bandung        -0.0488
location_Astanaanyar, Bandung         -0.0502
location_Gede Bage, Bandung           -0.0599
location_Buah Batu, Bandung           -0.0665
location_Bandung Kulon, Bandung       -0.0760
location_Antapani, Bandung            -0.0782
building_to_land_ratio                -0.0907
location_Rancasari, Bandung           -0.0956
location_Panyileukan, Bandung         -0.0966
location_Ujungberung, Bandung         -0.0995
location_Mandalajati, Bandung         -0.0999
location_Arcamanik, Bandung           -0.1131
location_Cibiru, Bandung              -0.1180
Name: price, dtype: float64
```

Gambar 5. Analisis Korelasi Terhadap Price

B. Pengembangan dan Evaluasi Model

Pada tahap modelling dataset dibagi menjadi dua yaitu data training dan juga data uji adapun dalam penelitian ini menggunakan perbandingan 80:20 atau 80% untuk data training dan 20% untuk data testing. Setelah itu dilakukan standarisasi menggunakan *standarscaler*. Standarisasi atau normalisasi data bisa mengurangi variabilitas antar fitur dan membuat model lebih mudah untuk belajar hubungan antara fitur dan target.

1) *Hyper Parameter Tuning*: Hyperparameter tuning adalah proses mengoptimalkan hyperparameter model untuk meningkatkan kinerja model pada data yang diberikan. Pada penelitian ini kami menggunakan metode *RandomizedSearchCV* dalam melakukan *hyperparameter tuning*. Dan didapatkan hasil:

TABEL 1.
BEST PARAMETER SETIAP MODEL

Model	Best Parameter
Random Forest Regressor	{'n_estimators': 400, 'min_samples_split': 5, 'min_samples_leaf': 1, 'max_features': 'log2', 'max_depth': 30, 'bootstrap': False}
Gradient Boosting Regressor	{'subsample': 0.8, 'n_estimators': 500, 'min_samples_split': 20, 'min_samples_leaf': 1, 'max_features': 'sqrt', 'max_depth': 5, 'learning_rate': 0.05}
XGBoost Regressor	{'subsample': 0.6, 'reg_lambda': 0, 'reg_alpha': 0, 'n_estimators': 400, 'min_child_weight': 5, 'max_depth': 5, 'learning_rate': 0.05, 'gamma': 0, 'colsample_bytree': 0.9}

C. Studi Komparatif

Untuk mengevaluasi kinerja masing-masing model diukur menggunakan metrik *Mean Absolute Error* (MAE) dan *R-squared* (R^2). Adapun hasil studi komparatif evaluasi kinerja model disajikan dalam tabel berikut:

Eksperimen 1 (menggunakan data *location* yang di encode) :

TABEL 1.
STUDI KOMPARATIF EVALUASI SETIAP MODEL EKSPERIMEN 1

No	Model	R2 Score	RMSE	MAE
1	*Stacking* Ensemble	90,76%	0,311	0,216
2	Random Forest Regressor	90,57%	0,314	0,212
3	*XGBoost Regressor*	90,14%	0,321	0,227
4	Gradient Boosting Regressor	89,73%	0,328	0,242
5	*Linear Regression*	86,32%	0,378	0,287

Eksperimen 2 (tidak menggunakan atribut *location*):

TABEL 2.
STUDI KOMPARATIF EVALUASI SETIAP MODEL EKSPERIMEN 2

No	Model	R2 Score	RMSE	MAE
1	*Stacking* Ensemble	85,84%	0,385	0,271
2	Random Forest Regressor	85,54%	0,389	0,263
3	*XGBoost Regressor*	85,11%	0,395	0,282
4	Gradient Boosting Regressor	84,52%	0,402	0,295
5	*Linear Regression*	80,15%	0,456	0,346

Dari hasil di atas menunjukkan bahwa kinerja model lebih baik pada eksperimen pertama dibandingkan dengan eksperimen kedua. Sehingga model lebih baik Ketika menggunakan atribut location dalam model prediksinya. Dan juga terlihat bahwa metode *Stacking* Ensemble memberikan kinerja terbaik dengan Score R2 mencapai 0.9076 , RMSE (*Root Mean Squared Erorr*) 0.3111 dan MAE (*Mean Average Error*) 0.2169.

D. Analisis Interpretabilitas

Hasil dari Analisis interpretabilitas model menggunakan SHAP (*SHapley Additive exPlanations*) values menunjukkan kontribusi masing-masing fitur terhadap prediksi harga rumah. Grafik SHAP pada Gambar 6 memperlihatkan bahwa land_area (luas tanah) memiliki pengaruh terbesar pada harga rumah, diikuti oleh building_area (luas bangunan) di posisi kedua.

Gambar 6. Analisis Interpretabilitas

Dari hasil analisis interpretabilitas, dapat diidentifikasi bagaimana fitur-fitur tersebut mempengaruhi prediksi harga:

Luas Tanah (*land_area*): Fitur ini konsisten menjadi prediktor terkuat dengan nilai SHAP positif yang dominan. Mengindikasikan bahwa setiap penambahan luas tanah memberikan kenaikan substansial pada harga prediksi. Dari sampel yang dianalisis, peningkatan luas tanah dari 100m² menjadi 200m² dapat meningkatkan prediksi harga hingga 80-90%, seperti terlihat pada kasus uji coba 1 dan 2 di Tabel 4.

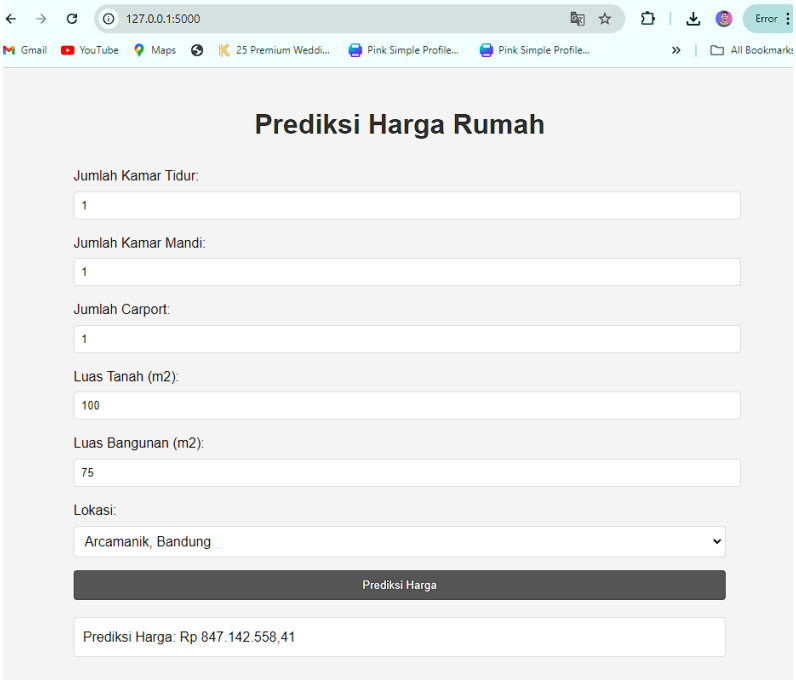
Luas Bangunan (*building_area*): Sebagai prediktor kedua terkuat, luas bangunan memiliki korelasi positif yang signifikan dengan harga rumah. Menariknya, pengaruh luas bangunan lebih kuat ketika kombinasinya dengan luas tanah menghasilkan rasio yang optimal. Pada uji coba 2 dan 3 di Tabel 4, dengan luas tanah yang sama (200m²), peningkatan luas bangunan dari 75m² menjadi 150m² meningkatkan prediksi harga sebesar 16.5%, menunjukkan bahwa meski penting, pengaruhnya tidak selinear luas tanah.

Jumlah Kamar dan Fasilitas: Fitur seperti *bedroom_count*, *bathroom_count*, dan *carport_count* memiliki pengaruh positif terhadap harga, tetapi dengan efek yang lebih kecil dibandingkan luas tanah dan luas bangunan. Pada uji coba 4 dan 5 di Tabel 4, peningkatan jumlah kamar tidur dari 1 ke 3, kamar mandi dari 1 ke 3, dan carport dari 1 ke 2 meningkatkan prediksi harga sekitar 4%, menunjukkan bahwa fitur-fitur ini memberikan nilai tambah yang lebih marginal dibandingkan luas tanah dan luas bangunan.

Lokasi (*location*): Faktor lokasi memiliki pengaruh yang sangat signifikan, dengan beberapa area premium seperti Bandung Wetan, Cibiru dan Sumurbandung yang mendapat nilai SHAP cukup tinggi. Seperti Sumurbandung, perbandingan antara uji coba 3 dan 4 di Tabel 4 menunjukkan bahwa dengan karakteristik fisik yang identik, perubahan lokasi dari Antapani ke Sumurbandung meningkatkan prediksi harga hingga 60%, menggarisbawahi premium lokasi yang substansial di daerah-daerah tertentu Bandung.

E. Integrasi Ke Website

Sebelum melakukan integrasi ke website Model di simpan dulu menggunakan joblib, agar bisa diimplementasikan atau diintegrasikan ke website. Lalu setelah di simpan lalu file model joblib tersebut di download, untuk di load di website.



Gambar 7. Tampilan Website

Gambar 7. merupakan hasil dari integrasi website menggunakan server flask yang mana diperlihatkan model *Stacking* ensemble bisa memprediksi harga rumah berdasarkan input yang diberikan oleh user.

TABEL 3.
UJI COBA PREDIKSI DI WEBSITE

Uji Coba	Kamar Tidur	Kamar Mandi	Carport	Luas Tanah (m2)	Luas Bangunan(m2)	Lokasi	Prediksi Harga
1	1	1	1	100	75	Antapani, Bandung	Rp 901.269.159
2	1	1	1	200	75	Antapani, Bandung	Rp 1.726.983.573
3	1	1	1	200	150	Antapani, Bandung	Rp 2.013.315.847
4	1	1	1	200	150	Sumurbandung, Bandung	Rp 3.224.109.105
5	3	3	2	200	150	Sumurbandung, Bandung	Rp 3.356.946.835

Tabel 4. adalah bentuk hasil uji coba sebanyak 5 kali percobaan dalam memprediksi harga menggunakan model *Stacking* ensemble yang telah diintegrasikan ke website menggunakan server flask.

IV. SIMPULAN

Berdasarkan hasil penilaian dan kajian, metode *Stacking Ensemble* terbukti sebagai pendekatan yang paling berhasil dalam meramalkan harga properti di Bandung pada tahun 2024, dengan hasil evaluasi terbaik yaitu R^2 sebesar 0,9076, RMSE sebesar 0,3111, dan MAE sebesar 0,2169, yang secara konsisten lebih unggul dibandingkan dengan model-model tunggal seperti *Random Forest, XGBoost*, dan *Gradient Boosting*. Keunggulan ini semakin terlihat pada percobaan yang memanfaatkan fitur lokasi yang telah di-encode, menegaskan pentingnya data spasial dalam meningkatkan ketepatan model. Kajian interpretabilitas menggunakan SHAP juga mengungkapkan bahwa fitur luas tanah (*land_area*) dan luas bangunan (*building_area*) menjadi faktor paling dominan dalam menentukan harga prediksi, di mana semakin tinggi nilai dari kedua fitur ini, maka harga yang diprediksi juga akan semakin meningkat.

Berdasarkan temuan ini, beberapa rekomendasi dapat diberikan kepada pemangku kepentingan. Bagi pengembang properti, optimalisasi rasio luas tanah dan bangunan serta prioritas pada area premium seperti Bandungwetan, Cibiru dan Sumurbandung dapat memaksimalkan nilai jual. Investor properti dapat memanfaatkan model ini untuk menganalisis potensi ROI dan diversifikasi portofolio di berbagai lokasi Bandung. Sementara pembuat kebijakan dapat mengimplementasikan tata ruang yang mempertimbangkan nilai properti berdasarkan atribut utama seperti luas tanah dan lokasi. Untuk penelitian selanjutnya, disarankan untuk mengeksplorasi teknik ensemble learning yang lebih canggih seperti CatBoost atau LightGBM, mengintegrasikan faktor makroekonomi, memperdalam analisis spasial dengan mempertimbangkan jarak ke fasilitas publik, mengembangkan model prediksi temporal jangka panjang (3-5 tahun), serta melakukan studi komparatif dengan kota-kota berkembang lainnya di Indonesia seperti Yogjakarta untuk mengidentifikasi faktor universal versus faktor lokal yang mempengaruhi harga properti.

UCAPAN TERIMAKASIH

Peneliti mengucapkan terima kasih yang sebesar-besarnya kepada STMIK AMIKOM Surakarta yang telah memberikan sumber daya dalam upaya penyelesaian penelitian ini.

DAFTAR PUSTAKA

[1] E. Febrion Rahayuningtyas, F. Novia Rahayu, Y. Azhar, and I. Artikel, "Prediksi Harga Rumah Menggunakan General Regression Neural Network," JURNAL INFORMATIKA, vol. 8, no. 1, 2021, doi: 10.31294/ji.v8i1.9036.

[2] M. Labib Mu'tashim, S. A. Damayanti, H. N. Zaki, T. Muhayat, and R. Wirawan, "Analisis Prediksi Harga Rumah Sesuai Spesifikasi Menggunakan Multiple *Linear Regression*," vol. 3, p. 2021, doi: 10.52166/j-macc.v4i1.2406.

[3] R. Mahendra Sanusi, A. Siswo, R. Ansori, R. Wijaya, and S. Si, "Prediksi Harga Rumah Di Kota Bandung Bagian Timur Dengan Menggunakan Metode Regresi Prediction Of House Prices In The East Bandung City Using The Regression Method."

[4] A. Fadilah, A. Siswo, R. Ansori, R. Wijaya, and S. Si, "Prediksi Harga Rumah Di Kota Bandung Bagian Timur Dengan Menggunakan Metode Moving Average Home Price Prediction In The East Bandung City With Moving Average."

[5] G. Najla, A. #1, and D. Fitrianah, "Penerapan Metode Regresi Linear Untuk Prediksi Penjualan Properti pada PT XYZ," Jurnal Telematika, vol. 14, no. 2, doi: 10.61769/telematika.v14i2.321.

[6] M. Radhi, D. Ryan Hamonangan Sitompul, S. Hamonangan Sinurat, and E. Indra, "Prediksi Harga Mobil Menggunakan Algoritma Regressi Dengan Hyper-Parameter Tuning," *Jurnal Sistem Informasi dan Ilmu Komputer Prima*, vol. 4, no. 2, 2021.
[7] N. Hadi and J. Benedict, "IMPLEMENTASI *MACHINE LEARNING* UNTUK PREDIKSI HARGA RUMAH MENGGUNAKAN ALGORITMA RANDOM FOREST," 2024. [Online]. Available: https://www.kaggle.com/harlfoxem/housesalesprediction
[8] I. D. Mienye and Y. Sun, "A Survey of Ensemble Learning: Concepts, Algorithms, Applications, and Prospects," 2022, *Institute of Electrical and Electronics Engineers Inc.* doi: 10.1109/ACCESS.2022.3207287.
[9] M. Mamun, A. Farjana, M. Al Mamun, and M. S. Ahammed, "Lung cancer prediction model using ensemble learning techniques and a systematic review analysis," in *2022 IEEE World AI IoT Congress, AIIoT 2022*, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 187–193. doi: 10.1109/AIIoT54504.2022.9817326.
[10] U. Indahyanti, N. L. Azizah, and H. Setiawan, "Pendekatan Ensemble Learning Untuk Meningkatkan Akurasi Prediksi Kinerja Akademik Mahasiswa," *Jurnal Sains dan Informatika*, vol. 8, no. 2, Dec. 2022, doi: 10.34128/jsi.v8i2.459.
[11] M. Sadikin and F. Alfiandi, "Comparative Study of Classification Method on Customer Candidate Data to Predict its Potential Risk," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 8, no. 6, p. 4763, Dec. 2018, doi: 10.11591/ijece.v8i6.pp4763-4771.
[12] J. Hong, H. Choi, and W. S. Kim, "A house price valuation based on the random forest approach: The mass appraisal of residential property in south korea," *International Journal of Strategic Property Management*, vol. 24, no. 3, pp. 140–152, Mar. 2020, doi: 10.3846/ijspm.2020.11544.
[13] E. Fitri, "Analisis Perbandingan Metode Regresi Linier, Random Forest Regression dan Gradient Boosted Trees Regression Method untuk Prediksi Harga Rumah," *JOURNAL OF APPLIED COMPUTER SCIENCE AND TECHNOLOGY (JACOST)*, vol. 4, no. 1, pp. 2723–1453, 2023, doi: 10.52158/jacost.491.
[14] Y. Meng, N. Yang, Z. Qian, and G. Zhang, "What makes an online review more helpful: An interpretation framework using xgboost and shap values," *Journal of Theoretical and Applied Electronic Commerce Research*, vol. 16, no. 3, pp. 466–490, 2021, doi: 10.3390/jtaer16030029.
[15] S. M. Lundberg, G. G. Erion, and S.-I. Lee, "Consistent Individualized Feature Attribution for Tree Ensembles," Feb. 2018, doi: 10.48550/arXiv.1802.03888 [Online]. Available: http://arxiv.org/abs/1802.03888