Implementasi Algoritma *Random Forest Regression* dalam Sistem Prediksi Harga Rumah di Jabodetabek

I Made Gede Aryadana Baraja Putra[a1], I Ketut Gede Suhartana[a2]

[a]Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Udayana
Jalan Raya Kampus Udayana, Bukit Jimbaran, Kuta Selatan, Badung, Bali, Indonesia
[1]putra.2308561070@student.unud.ac.id
[2]ikg.suhartana@unud.ac.id

Abstract

Indonesia's rapid urbanization, particularly in the Jabodetabek region, has created a severe housing shortage with a backlog of 2.93 million units representing 30% of the national deficit. This imbalance between supply and demand necessitates accurate house price prediction systems to guide market participants. This research implements Random Forest Regression algorithm to predict house prices in the Jabodetabek region using comprehensive datasets covering land area, building area, geographical location, room quantities, facilities, and property characteristics across districts and cities. The methodology involves data preprocessing, model training using Random Forest Regression, and performance evaluation using established metrics. Results demonstrate great algorithm performance with RMSE of 0.3545, MAE of 0.2014, MAPE of 1.0184, and R^2 of 0.8751 confirming the model explains 87.51% of house price variance. The implementation successfully addresses the research objective of providing developers with a reliable algorithmic framework for property pricing strategies.

Keywords: *Random Forest Regression, House Price Prediction, Jabodetabek, Machine Learning, Property Valuation*

1. Pendahuluan

Urbanisasi yang terjadi di Indonesia kian meningkat. Berdasarkan data dari Badan Pusat Statistik pada tahun 2023, lebih dari 56% masyarakat Indonesia telah menetap di daerah perkotaan [1]. Studi menunjukkan bahwa tingkat urbanisasi yang tinggi memicu lonjakan penduduk kota dan permintaan hunian [2]. Jabodetabek menjadi salah satu wilayah perkotaan terbesar yang memiliki tingkat urbanisasi yang tinggi. Peluang kerja dan infrastruktur yang memadai mengakibatkan jutaan penduduk berpindah ke wilayah Jabodetabek.

Sayangnya, peningkatan jumlah penduduk tidak diimbangi dengan peningkatan jumlah pembangunan rumah yang cukup. Data PUPR (HREIS) pada tahun 2021 menunjukkan *backlog* perumahan Jabodetabek sekitar 2,93 juta unit atau 30% dari total *backlog* perumahan nasional pada tahun 2022 hingga 2023 [3]. Kurangnya jumlah pasokan rumah memicu peningkatan harga rumah sehingga meningkatkan persaingan antar penduduk untuk memperoleh rumah layak. Oleh karena itu, diperlukan wawasan bagi masyarakat mengenai faktor penentu harga jual rumah agar dapat bersaing dalam memilih rumah yang terjangkau. Permintaan, lokasi dan karakteristik rumah, seperti luas lahan, bangunan, serta jumlah ruangan turut menjadi faktor penentu harga jual. Beberapa faktor lainnya turut mempengaruhi harga jual rumah, seperti kondisi kelistrikan, perabotan, dan fasilitas penunjang lainnya [4].

Kompleksitas penentuan harga jual rumah mendorong berbagai peneliti untuk mengembangkan pendekatan prediktif berbasis algoritma pembelajaran mesin. Terdapat beberapa algoritma pembelajaran mesin yang cocok digunakan dalam pemodelan sistem prediksi, seperti algoritma *Random Forest Regression* dan *Multiple Linear Regression* yang digunakan untuk memprediksi harga rumah di Kawasan *Elite* Jabodetabek [5]. Dari penelitian tersebut dapat diperoleh hasil bahwa model dengan algoritma *Random Forest Regression* menggunakan 10 variabel bebas

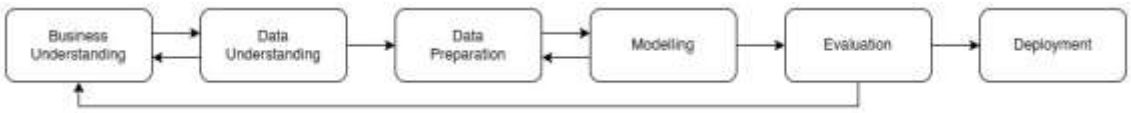
p-ISSN: 2986-3929
e-ISSN: 3032-1948

merupakan model teroptimal dengan nilai R^2 sebesar 0,8465. Sementara itu, terdapat penelitian yang juga membandingkan algoritma *Random Forest Regression* dengan *Linear Regression* dan *Gradient Boosted Trees Regression* untuk memprediksi harga rumah di daerah Tebet dan Jakarta Selatan [6]. Hasil penelitian tersebut menunjukkan bahwa algoritma *Random Forest Regression* memperoleh akurasi tertinggi yaitu sebesar 81,5% dengan nilai RMSE sebesar 0,440.

Berdasarkan tinjauan literatur tersebut, *Random Forest Regression* menunjukkan kinerja yang baik dalam memprediksi harga rumah. Oleh karena itu, penelitian ini akan mengimplementasikan algoritma *Random Forest Regression* dalam melakukan prediksi terhadap harga jual rumah di kawasan Jabodetabek. Penelitian ini berfokus pada *dataset* harga jual rumah di seluruh wilayah Jabodetabek dengan variabel yang digunakan untuk memprediksi harga rumah, diantaranya luas tanah, luas bangunan, lokasi geografis, jumlah kamar tidur dan kamar mandi, kondisi kelistrikan, dan perabotan atau fasilitas penunjang lainnya. Penelitian ini bertujuan sebagai acuan untuk pengembang agar dapat memprediksi harga rumah dengan algoritma yang sesuai sehingga hasil prediksi harga rumah dapat menggambarkan estimasi harga jual rumah pada kondisi lapangan.

2. Metode Penelitian

Penelitian ini dilakukan berdasarkan tahapan dari kerangka kerja CRISP-DM (*Cross-Industry Standard Process for Data Mining*). CRISP-DM merupakan standar proses analisis data yang bertujuan untuk memperoleh pemahaman dari data yang digunakan untuk memperoleh solusi dalam permasalahan bisnis. CRISP-DM terdiri dari 6 tahapan, yaitu *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modelling*, *Evaluation*, dan *Deployment* [5], [7].



Gambar 1. Kerangka Kerja CRISP-DM

2.1. *Business Understanding*

Business understanding merupakan tahap awal dimana permasalahan harus diidentifikasi untuk mendapatkan gambaran umum tentang sumber daya yang dibutuhkan dan tersedia. Tahapan ini juga menjelaskan tujuan, jenis, serta kriteria keberhasilan penambangan data [7].

2.2. *Data Understanding*

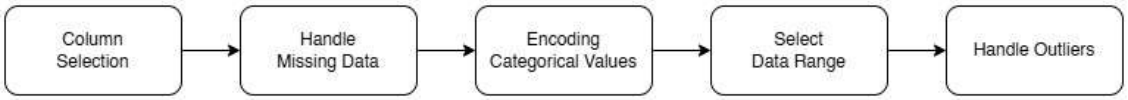
Tahap *data understanding* bertujuan untuk memperoleh pemahaman lebih lanjut terhadap data yang digunakan sehingga hal-hal yang harus dilakukan pada tahap selanjutnya dapat ditentukan berdasarkan pemahaman pada tahap ini. Tahap ini melibatkan pengumpulan data dari sumber data, eksplorasi, pemeriksaan kualitas data, dan deskripsi data yang menggunakan analisis statistik untuk menentukan karakteristik dan kolasi dari setiap atribut pada data [5], [7].

Pada penelitian ini, dataset yang dikumpulkan adalah dataset *open source* pada platform Kaggle yang berisi hasil *web scrapping* terhadap rumah di Jabodetabek dari situs web rumah123.com. Dataset terdiri dari 3.553 *record* dan 27 kolom, dengan rincian yaitu 13 kolom bertipe numerikal dan 14 kolom bertipe kategorikal. Dataset akan dieksplorasi lebih lanjut untuk dapat menentukan tahap preparasi data yang sesuai.

2.3. *Data Preparation*

Tahap *data preparation* bertujuan untuk mempersiapkan data sesuai dengan kriteria inklusi dan eksklusi dari model yang digunakan. Tahap ini meliputi pemilihan data dan atribut, pembersihan data, serta pemeriksaan seluruh nilai pada setiap atribut yang digunakan untuk mencegah ketidakseimbangan pada data [7]. Rincian tahap *data preparation* dapat dilihat pada gambar 2.

p-ISSN: 2986-3929
e-ISSN: 3032-1948



Gambar 2. Tahap *Data Preparation*

2.4. *Modelling*

Tahap *modelling* terdiri dari pemilihan algoritma, pemisahan data untuk membuat kasus latih dan uji, serta pelatihan dan pengujian model [7]. Penelitian ini menggunakan algoritma *Random Forest Regression* dengan *Grid Search Cross Validation* sebagai proses *hyperparameter tuning* untuk mendapat parameter terbaik pada model.

a *Random Forest Regression*

Random Forest Regression adalah algoritma *ensemble learning* berbasis pohon keputusan (*decision trees*) yang digunakan dalam kasus regresi. Secara intuitif, algoritma ini membangun banyak pohon regresi (*regression tree*) secara acak dan menggabungkan prediksinya dengan rata-rata untuk menghasilkan prediksi akhir yang lebih akurat dan stabil [8]. Dalam prosesnya, *Random Forest* memanfaatkan metode *bagging* (*bootstrap aggregating*) dan pemilihan subruang fitur (*random subspace*) untuk mengurangi varians model dan mencegah *overfitting* [9].

Dalam regresi Random Forest, prediksi akhir untuk suatu observasi x adalah rata-rata dari prediksi seluruh pohon dalam hutan. Jika terdapat B pohon dengan fungsi prediksi $f_b(x)$, maka:

$$\hat{y}(x) = \frac{1}{B}\sum_{b=1}^{B} f_b(x) \quad (1)$$

Random Forest Regression terdiri dari beberapa tahapan, yakni sebagai berikut.

- *Bootstrapping* Data
  Dari dataset latih, dibuat B sampel baru secara acak dengan penggantian (*bootstrap*). Masing-masing sampel latih akan menghasilkan satu pohon. Sekitar sepertiga sampel yang tidak digunakan dalam proses *bootstrapping* dapat digunakan untuk validasi model.
- Membangun Pohon Keputusan
  Untuk setiap *bootstrap sample*, dibuat pohon keputusan regresi dengan proses CART (*Classification and Regression Tree*). Pada setiap simpul, pohon mencari pembagian (*split*) terbaik berdasarkan subset acak dari fitur (parameter *max_features*). Kriteria *split* biasanya menggunakan MSE sehingga setiap split meminimalkan variansi target di tiap cabang.
- Prediksi per Pohon
  Setelah pohon terlatih, diberikan satu observasi input x pada setiap pohon. Setiap pohon menghasilkan prediksi $f_b(x)$ sesuai dengan nilai *output* di daun (*leaf*) tempat x berakhir.
- Agregasi Hasil
  Seluruh prediksi pohon dikumpulkan untuk dihitung rata-ratanya. Pada kasus regresi, prediksi akhir model adalah rata-rata prediksi seluruh pohon.

b *Grid Search Cross Validation*

Grid Search adalah metode eksplorasi menyeluruh terhadap kombinasi *hyperparameter* yang telah ditentukan secara manual. *Cross Validation* adalah metode *resampling* untuk mengestimasi performa model secara *robust*. *Grid Search Cross Validation*

menggabungkan konsep *Grid Search* dan *Cross Validation*, yakni pengujian seluruh kombinasi *hyperparameter* untuk setiap data *training* dan *testing* yang dihasilkan oleh proses *cross validation* [10].

2.5. *Evaluation*

Model yang telah dilatih dan diuji pada tahap *modelling* akan melalui tahap evaluasi melalui pengukuran berbagai metrik. Penelitian ini akan berfokus pada empat metrik evaluasi, yaitu *Mean Absolute Error* (MAE), *Mean Absolute Percentage Error* (MAPE), *Root Mean Squared Error* (RMSE), dan R-*Squared* (R^2). Jika terdapat kumpulan nilai aktual y dan nilai prediksi $\hat{y}$ sejumlah n, maka secara matematis MAE, MAPE, RMSE, dan R^2 dapat dirumuskan sebagai berikut.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (2)$$

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \quad (3)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (4)$$

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

2.6. *Deployment*

Deployment merupakan tahap akhir dari CRISP-DM. Tahap *deployment* bertujuan untuk menyebarluaskan hasil penelitian agar dapat digunakan oleh pengguna umum. *Deployment* pada penelitian ini berupa pengembangan web menggunakan *framework* Streamlit berbasis bahasa pemrograman Python untuk mempermudah integrasi dengan model yang diperoleh.

3. Hasil dan Diskusi

3.1. *Business Understanding*

Penelitian ini berfokus pada untuk prediksi harga rumah di kawasan Jabodetabek menggunakan algoritma berjenis regresi. Melalui penelitian ini, diharapkan masyarakat khususnya di wilayah Jabodetabek dapat memperkirakan harga rumah yang ingin dibeli sesuai dengan spesifikasi yang diinginkan. Selain itu, diharapkan juga masyarakat umum dapat memperoleh wawasan mengenai persebaran harga rumah di wilayah Jabodetabek sehingga meminimalisir kerugian yang ditimbulkan ketika melakukan transaksi jual beli rumah.

3.2. *Data Understanding*

a Pengumpulan Data

Data yang digunakan pada penelitian adalah data berjenis sekunder. Dataset diperoleh dari situs web kaggle.com yang dapat diakses melalui [11]. Dataset tersebut merupakan data yang diperoleh melalui proses *web scrapping* pada situs web rumah123.com dan disimpan dalam format .csv. Dataset terdiri dari 3.553 *record* dan 27 kolom, meliputi *url, price_in_rp, title, address, district, city, lat, long, facilities, property_type, ads_id, bedrooms, bathrooms, land_size_m2, building_size_m2, carports, certificate, electricity, maid_bedrooms, maid_bathrooms, floors, building_age, year_built, property_condition,*

p-ISSN: 2986-3929
e-ISSN: 3032-1948

building_orientation, garages, dan *furnishing*. Berikut merupakan penjelasan dari setiap kolom pada dataset.

Tabel 1. Kolom pada Dataset

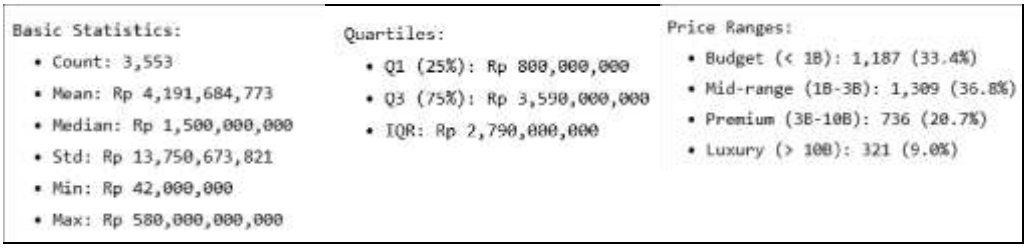
Atribut	Keterangan	Nilai Unik
url	Tautan halaman penjualan rumah	3552
price_in_rp	Harga penjualan rumah	660
title	Judul rumah yang dijual pada situs web	3342
address	Keterangan alamat rumah yang dijual	397
district	Kecamatan alamat rumah yang dijual	380
city	Kota alamat rumah yang dijual	9
lat	Nilai latitude atau lintang sesuai dengan titik pusat district	389
long	Nilai longitude atau bujur sesuai dengan titik pusat district	390
facilities	Daftar fasilitas yang disediakan ketika membeli rumah	2024
property_type	Tipe properti yang dijual (seluruh data memiliki nilai ‘rumah’)	1
ads_id	Id ads unik (identifier)	3457
bedrooms	Jumlah kamar tidur yang ada di rumah	22
bathrooms	Jumlah kamar mandi yang ada di rumah	22
land_size_m2	Luas lahan keseluruhan	481
building_size_m2	Luas bangunan keseluruhan	358
carports	Jumlah carports yang tersedia di rumah	13
certificate	Jenis sertifikat yang dilampirkan saat penjualan	4
electricity	Kapasitas tegangan listrik di rumah	30
maid_bedrooms	Jumlah kamar tidur khusus ART	8
maid_bathrooms	Jumlah kamar mandi khusus ART	6
floors	Jumlah lantai/tingkat di rumah	5
building_age	Usia rumah terhitung sejak dibangun	42
year_built	Tahun dibangunnya rumah	46
property_condition	Kondisi singkat properti rumah	7
building_orientation	Orientasi/arah bangun rumah	8
garages	Jumlah garasi di rumah	11
furnishing	Kondisi perabotan/perlengkapan di rumah	4

b Analisis Data Eksploratif

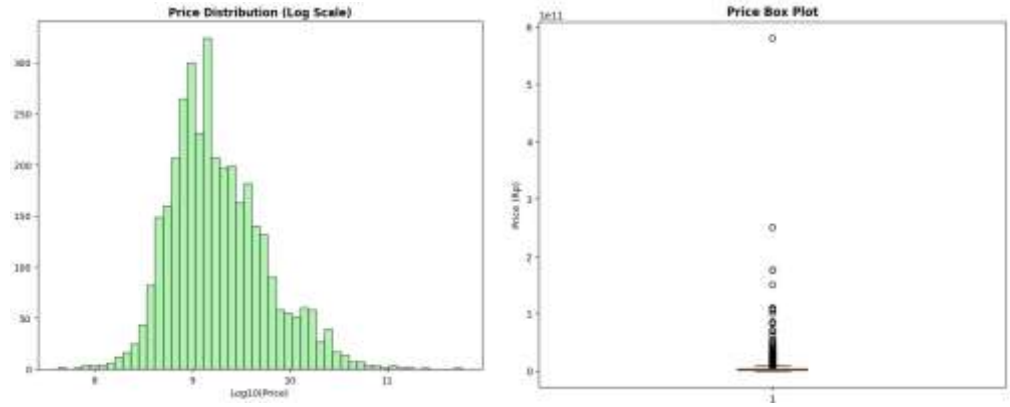
Analisis Data Eksploratif bertujuan untuk memperoleh pemahaman lebih lanjut mengenai dataset yang digunakan. Pada penelitian ini, analisis data eksploratif yang dilakukan terdiri dari beberapa tahapan, diantaranya sebagai berikut. Pada tabel 1, terdapat kolom jumlah nilai unik yang dimiliki oleh setiap kolom pada dataset. Kolom *url, title,* dan *ads_id* memiliki jumlah nilai unik yang hampir setara dengan jumlah data, yaitu 3.553 *record*. Dapat

p-ISSN: 2986-3929
e-ISSN: 3032-1948

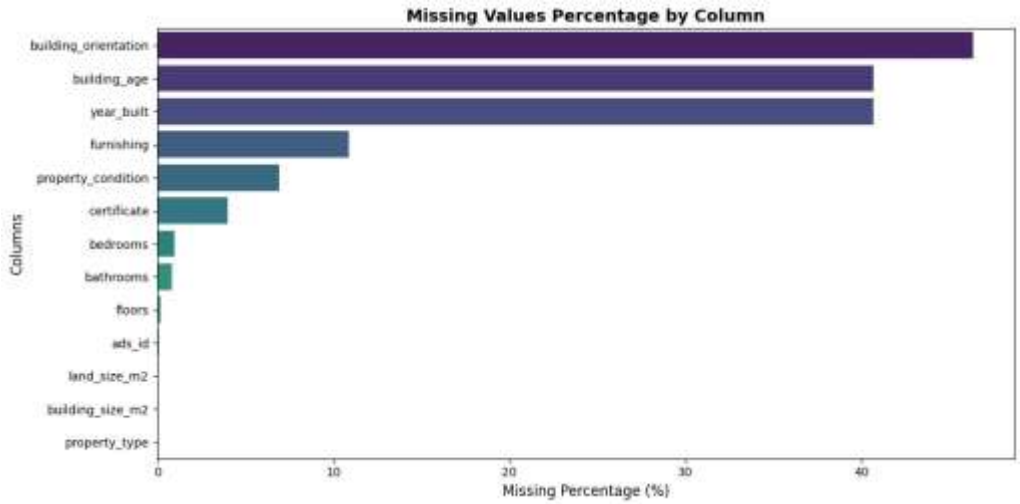
dikatakan bahwa ketiga kolom tersebut memiliki nilai yang berbeda antara *record* satu dengan lainnya, dan ketiga kolom tersebut menjadi *identifier* dari masing-masing *record*. Hal sebaliknya terjadi pada kolom *property_type* dimana kolom ini hanya memiliki satu nilai yang sama untuk seluruh *record*, yaitu 'rumah'. Gambar 3 menjelaskan analisis statistik dari kolom *price_in_rp* pada dataset. Terdapat beberapa perhitungan statistik dasar, seperti *mean, median, std, min, max,* Q1, Q3, dan IQR. Dari gambar 3 dan 4, dapat dilihat bahwa mayoritas data tersebar merata dengan nilai kuantil 1 adalah 800 juta rupiah dan nilai kuantil 3 adalah 3,59 miliar rupiah. Namun, terdapat beberapa data *outlier* yang sangat jauh dari persebaran data. Hal tersebut dibuktikan dengan adanya nilai maksimum yaitu 580 miliar rupiah dan beberapa nilai lainnya yang tergolong sebagai bagian dari *Luxury*.



Gambar 3. Analisis Variabel Harga (Dependen)



Gambar 4. Distribusi Data Harga



Gambar 5. Analisis Nilai Kosong Berdasarkan Kolom

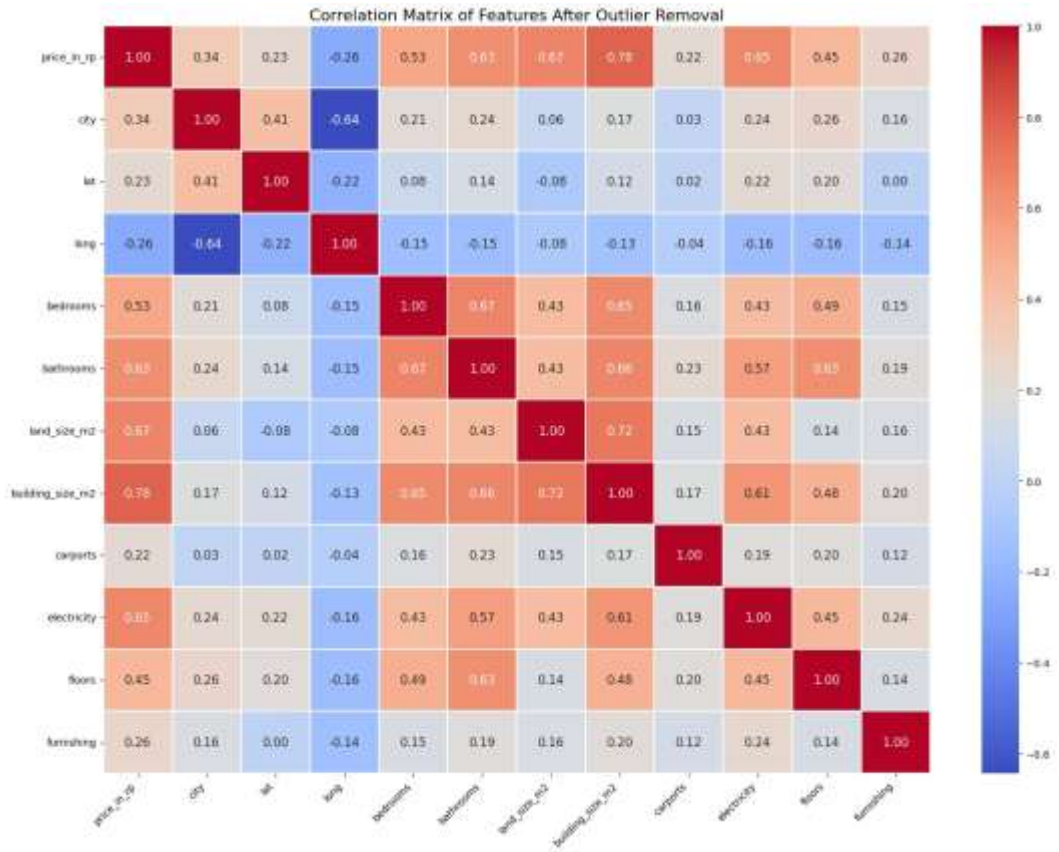
p-ISSN: 2986-3929
e-ISSN: 3032-1948

Gambar 5 merupakan grafik jumlah nilai kosong (*missing value*) dari setiap kolom. Dapat dilihat bahwa terdapat tiga kolom yang memiliki nilai kosong di atas 40% dari jumlah data, yaitu *building_orientation, building_age,* dan *year_built*. Berdasarkan informasi tersebut, ketiga kolom tersebut tidak akan digunakan dalam perhitungan mengingat jumlah data kosong yang hampir mencapai setengah dari jumlah data dapat merusak proses pengolahan data.

3.3. *Data Preparation*

Setelah memperoleh pemahaman lebih lanjut mengenai dataset yang digunakan, maka selanjutnya adalah tahap *data preparation* untuk mempersiapkan data sebelum menjadi *input* dalam rancangan model yang dibangun. Penelitian ini menerapkan beberapa tahapan dalam *data preparation*, diantaranya sebagai berikut.

a. *Column Selection*
Pada tahap ini, kolom yang memiliki korelasi yang kuat dengan harga rumah akan digunakan dalam *modelling*. Sedangkan, kolom yang memiliki korelasi yang kurang kuat maupun memiliki nilai kosong yang banyak tidak akan dilibatkan dalam proses *modelling*. Hasil dari pemilihan kolom adalah 12 kolom, yaitu *price_in_rp, city, lat, long, bedrooms, bathrooms, land_size_m2, building_size_m2, carports, electricity, floors,* dan *furnishing*. *Heatmap correlations* dari kolom yang diseleksi dapat dilihat pada gambar 6.



Gambar 6. *Heatmap Correlations* Antar Kolom

b. *Handle Missing Data*
Tabel 2 merupakan persebaran nilai hilang atau *missing value* dari kolom terpilih. Untuk kolom dengan jumlah nilai hilang yang sedikit, seperti *bedrooms, bathrooms, floors, land_size_m2,* dan *building_size_m2, record* yang bersangkutan akan dihapus dari dataset yang digunakan. Pada kolom *furnishing*, nilai hilang diisi menggunakan imputasi data. Hasil akhir dari proses ini adalah 3.510 *record* pada 12 kolom.

p-ISSN: 2986-3929
e-ISSN: 3032-1948

Tabel 2. Persebaran Nilai Hilang dari Kolom Terpilih

Kolom Terpilih	Jumlah Nilai Hilang
bedrooms	34
bathrooms	29
land_size_m2	2
building_size_m2	2
floors	6
furnishing	387

c. *Encoding Categorical Values*
Pada dataset, terdapat tiga kolom kategorikal yang perlu dilakukan pengkodean, yaitu *city, electricity,* dan *furnishing*. Proses *encoding* pada kolom *city* dan *furnishing* dilakukan dengan mengubah setiap nilai unik menjadi angka dari 0 sampai n. Proses *encoding* pada kolom *electricity* dilakukan dengan mengambil nilai angka pada setiap nilai unik (contoh: '1300 mah' akan dikodekan menjadi 1300).

d. *Select Data Range*
Melihat persebaran data harga pada proses analisis data eksploratif, diputuskan untuk menggunakan data dengan nilai harga berada pada rentang 300 juta rupiah hingga 50 miliar rupiah. Hal tersebut dilakukan mengingat terdapat kategori rumah *budget* dan *luxury* yang memang terjadi di kondisi lapangan, sehingga data yang tidak memenuhi rentang tersebut tidak akan digunakan selama proses *modelling*. Hasil akhir dari proses ini adalah 3.415 *record* pada 12 kolom.

e. *Handle Outliers*
Proses memilih rentang data saja tidak cukup untuk membersihkan data karena rentang yang dipilih cukup besar sehingga beberapa *record* dapat dikategorikan sebagai *outlier*. Oleh karena itu, *outlier* perlu ditangani untuk mencegah kejanggalan pada proses *modelling*. Pada penelitian ini, *outlier* ditangani dengan metode rentang interkuartil. Hasil akhir dari proses ini adalah 2.716 *record* pada 12 kolom.

3.4. *Modelling*

Data yang sudah dipreparasi akan menjalani dua tahapan sebelum digunakan sebagai *input* dari model yang dirancang, yakni *splitting* dan *scaling*. Data dibagi menjadi dua bagian, yaitu X yang berperan sebagai variabel bebas dan y yang berperan sebagai variabel terikat, yaitu harga rumah. Kemudian, data melalui proses *train_test_split* untuk memperoleh data latih dan data uji dengan proporsi yakni 80:20. Penelitian ini menggunakan metode *Standard-Scaler* untuk *scaling* dataset yang telah dibagi menjadi data latih dan uji.

Model yang dirancang pada penelitian ini menggunakan algoritma *Random Forest Regression* dengan *Grid Search Cross Validation* sebagai pengujian *hyperparameter tuning* dan *cross validation*. Terdapat beberapa parameter yang diuji, yakni *n_estimator* dengan pilihan nilai yaitu [50, 100, 200], *max_depth* dengan pilihan nilai yaitu [10, 20, 30, None], *min_sample_split* dengan pilihan nilai yaitu [2, 5, 10], *mean_sample_leaf* dengan pilihan nilai yaitu [1, 2, 4], dan *max_features* dengan pilihan nilai yaitu ['sqrt', 'log2']. *Cross validation* dijalankan menggunakan 5 lipatan. Setelah konfigurasi selesai, model dilatih dan diuji untuk mengukur metrik evaluasi.

p-ISSN: 2986-3929
e-ISSN: 3032-1948

3.5. *Evaluation*

```
Starting GridSearchCV for Random Forest Regression...
This may take a few minutes...
Fitting 5 folds for each of 216 candidates, totalling 1080 fits

Best parameters: {'max_depth': 30, 'max_features': 'sqrt', 'min_samples_leaf': 1, 'min_samples_split': 2, 'n_estimators': 100}
Best cross-validation score: 0.2098
```

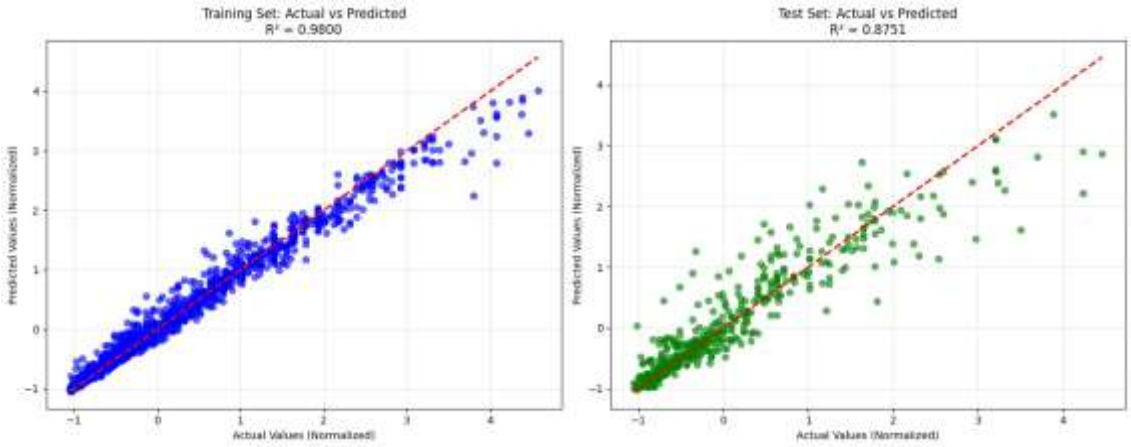
Gambar 7. Hasil Parameter Optimal Melalui GridSearchCV

Berdasarkan gambar 7, dapat diperoleh bahwa model memiliki kinerja yang optimal dengan parameter *max_depth* bernilai 30, *max_feature* bernilai 'sqrt', *min_samples_leaf* bernilai 1, *min_samples_split* bernilai 2, dan *n_estimator* bernilai 100.

Training Set Performance:	Test Set Performance:
MAE: 0.0769	MAE: 0.2014
RMSE: 0.1415	RMSE: 0.3545
R^2: 0.9800	R^2: 0.8751
MAPE: 0.3228	MAPE: 1.0814

Gambar 8. Hasil Pengukuran Metrik Evaluasi Model

Gambar 8 menunjukkan hasil pengukuran metrik evaluasi pada model yang diuji. Pada data *training*, diperoleh nilai MAE, MAPE, dan RMSE yang tergolong rendah, serta nilai R^2 sebesar 0,9800 atau setara dengan 98,00% akurasi. Pada data *testing*, diperoleh nilai MAE, MAPE, dan RMSE yang juga tergolong renda, serta nilai R^2 sebesar 0,8751 atau setara dengan 87,51% akurasi. Berdasarkan nilai galat dari MAE, MAPE, RMSE, dan nilai R^2, model yang telah melalui proses *hyperparameter tuning* dan *cross validation* memiliki kinerja yang baik. Visualisasi plot hasil prediksi dan aktual pada *training* dan *testing* dapat dilihat pada gambar 9.



Gambar 9. Plot Hasil Prediksi dengan Aktual

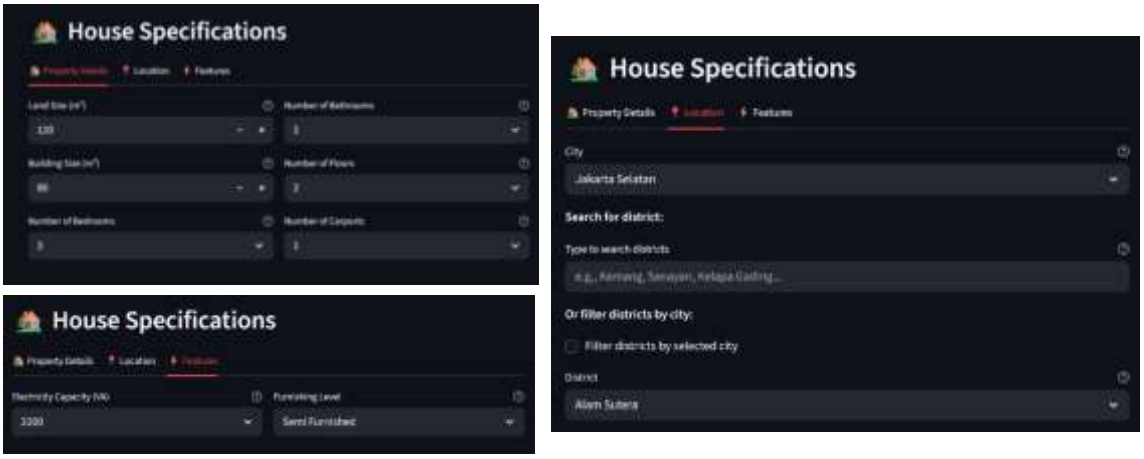
Selanjutnya, model dengan parameter terbaik akan diekspor menggunakan *library* pickle agar dapat digunakan pada proses penyebaran atau *deployment*. Model akan disimpan bersama dengan *parameter lainnya*, seperti *scaler* dan *metadata* untuk menjaga hasil pelatihan model.

3.6. *Deployment*

Tahap ini melibatkan model yang telah diekspor sebagai otak untuk melakukan prediksi di dalam aplikasi web prediksi rumah di Jabodetabek. Pada aplikasi web, pengguna disediakan berbagai

p-ISSN: 2986-3929
e-ISSN: 3032-1948

kolom terkait spesifikasi rumah yang perlu diisi. Kolom isian terbagi menjadi tiga kategori, yaitu spesifikasi atau detail rumah, lokasi, dan fitur kelistrikan. Rincian dari kolom isian pengguna dapat dilihat pada gambar 10.



Gambar 10. Tampilan Situs Web Kolom Isian Prediksi Harga Rumah di Jabodetabek

4. Kesimpulan

Berdasarkan hasil dari keseluruhan tahapan penelitian yang dilakukan, dapat disimpulkan bahwa model yang dirancang memiliki kinerja yang optimal melalui proses *hyperparameter tuning.* Hal ini ditunjukkan dengan perolehan parameter optimal, yaitu *max_depth* bernilai 30, *max_feature* bernilai 'sqrt', *min_samples_leaf* bernilai 1, *min_samples_split* bernilai 2, dan *n_estimator* bernilai 100. Dari model tersebut, diperoleh bahwa algoritma *Random Forest Regression* memiliki kinerja yang optimal dalam menangani kasus prediksi. Hal ini dibuktikan dengan nilai metrik evaluasi galat yang tergolong rendah, yaitu MAE berada di angka 0,2014, MAPE berada di angka 1,0184, dan RMSE berada di angka 0,3545, serta nilai R^2 yang tergolong tinggi, yaitu mencapai angka 0,8751 pada tahap pengujian. Hal ini sesuai dengan hasil penelitian [5], [6] yang menunjukkan bahwa algoritma *Random Forest Regression* mengungguli performa dari algoritma regresi lainnya. Secara keseluruhan, penelitian ini berhasil mencapai tujuan utama yang ditetapkan, yakni implementasi algoritma *Random Forest Regression* dalam sistem prediksi harga rumah di Jabodetabek serta dapat memberikan kontribusi yang signifikan untuk memberikan wawasan dan memudahkan masyarakat dalam memprediksi harga rumah sesuai dengan spesifikasinya.

Daftar Pustaka

[1] Badan Pusat Statistik, *Statistik Indonesia 2023*. Jakarta: Badan Pusat Statistik, 2023.
[2] N. F. Arsaf, Bakhtiar, and Ahmadin, "Dampak Urbanisasi terhadap Ketersediaan dan Keterjangkauan Perumahan di Kota Besar," *QISTINA: Jurnal Multidisiplin Indonesia*, vol. 4, no. 1, pp. 190–197, Jun. 2025.
[3] A. Adri and S. Ato, "Jabodetabek Masih Kekurangan 2,9 Juta Rumah," Kompas.id. Accessed: Jun. 26, 2025. [Online]. Available: https://www.kompas.id/baca/metro/2023/02/10/jabodetabek-masih-kekurangan-29-jutarumah
[4] A. A. G. S. Utama, "The Best Model and Variables Affecting Housing Values of Big Cities in Indonesia," *Galaxy International Interdisciplinary Research Journal (GIIRJ)*, vol. 10, no. 6, pp. 782–793, Jun. 2022, [Online]. Available: https://www.researchgate.net/publication/361466434
[5] N. A. C. Putri and D. B. Arianto, "Komparasi Penggunaan Information Gain Pada Machine Learning untuk Memprediksi Harga Rumah di Jabodetabek," *Jurnal Sains dan Teknologi*, vol. 5, no. 3, pp. 756–762, Feb. 2024, doi: 10.55338/saintek.v5i1.2052.
[6] 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. 58–64, 2023, doi: 10.52158/jacost.491.

[7] C. Schröer, F. Kruse, and J. M. Gómez, "A Systematic Literature Review on Applying CRISP-DM Process Model," in *Procedia Computer Science*, Elsevier B.V., 2021, pp. 526–534. doi: 10.1016/j.procs.2021.01.199.

[8] M. Mao, "A Comparative Study of Random Forest Regression for Predicting House Prices Using," *Highlights in Science, Engineering and Technology CSIC*, vol. 85, pp. 969–974, 2024.

[9] IBM, "Apa itu random forest?," IBM. Accessed: Jun. 29, 2025. [Online]. Available: https://www.ibm.com/id-id/think/topics/random-forest

[10] G. Malato, "Hyperparameter tuning. Grid search and random search," Your Data Teacher. Accessed: Jun. 29, 2025. [Online]. Available: https://www.yourdatateacher.com/2021/05/19/hyperparameter-tuning-grid-search-andrandom-search

[11] N. Barizki, "Daftar Harga Rumah Jabodetabek," Kaggle. Accessed: Jun. 29, 2025. [Online]. Available: https://www.kaggle.com/datasets/nafisbarizki/daftar-harga-rumahjabodetabek

p-ISSN: 2986-3929
e-ISSN: 3032-1948

Halaman ini sengaja dibiarkan kosong