

Perbandingan Model Regresi Linier dan Random Forest Regressor dalam Estimasi Harga Jual Rumah Berdasarkan Data Properti di Yogyakarta

Indriani Zabrina Putri^{1*}, Mila Rosyada², Salma Elsa Widyadhana³, Saskia Aila Virda⁴,
Muhammad Arifin⁵

¹Fakultas Teknik, Program Studi Sistem Informasi, Universitas Muria Kudus, Kudus, Indonesia

Email: ^{1*}indriputri12278@gmail.com, ²mlarosyada08@gmail.com, ³salmaelsa1874@gmail.com

⁴ailavirda7@gmail.com, ⁵arifin.m@umk.ac.id

Abstrak—Penentuan harga jual rumah merupakan aspek penting dalam transaksi properti, khususnya di wilayah dengan dinamika pasar yang tinggi seperti Yogyakarta. Penelitian ini membandingkan dua pendekatan pemodelan prediktif, yaitu Regresi Linier dan Random Forest Regressor, dalam mengestimasi harga jual rumah berdasarkan data properti yang diperoleh dari situs rumah123.com. Dataset yang digunakan terdiri dari 1.036 entri, mencakup variabel-variabel seperti harga, luas tanah, luas bangunan, jumlah kamar tidur, jumlah kamar mandi, ketersediaan carport, dan lokasi. Setelah melalui tahap pra-proses data, kedua model dilatih dan diuji menggunakan data yang sama untuk mengukur performa prediksi. Hasil evaluasi menunjukkan bahwa model Random Forest memiliki keunggulan dalam akurasi dibandingkan Regresi Linier, khususnya dalam menangani variasi data dan hubungan non-linier antar variabel. Meskipun Regresi Linier menghasilkan nilai koefisien determinasi (R^2) sebesar 0,846 yang menunjukkan kemampuan model dalam menjelaskan 84,6% variabilitas harga Random Forest menunjukkan prediksi yang lebih presisi pada data uji. Temuan ini menegaskan bahwa pemilihan model yang tepat sangat bergantung pada kompleksitas data dan kebutuhan akurasi. Studi ini memberikan kontribusi penting dalam pengembangan sistem pendukung keputusan berbasis data untuk estimasi harga properti dan menjadi dasar bagi penelitian lanjutan dengan pendekatan machine learning yang lebih kompleks.

Kata Kunci: Estimasi Harga Rumah; Regresi Linier; Random Forest; Prediksi Harga; Yogyakarta

Abstract—Determining the selling price of a house is a crucial aspect in property transactions, especially in regions with dynamic market conditions such as Yogyakarta. This study compares two predictive modeling approaches Linear Regression and Random Forest Regressor in estimating house prices based on property data obtained from the rumah123.com website. The dataset used consists of 1,036 entries, covering variables such as price, land area, building area, number of bedrooms, number of bathrooms, availability of a carport, and location. After undergoing data preprocessing, both models were trained and tested using the same dataset to assess their predictive performance. Evaluation results indicate that the Random Forest model outperforms Linear Regression in terms of accuracy, particularly in handling data variation and non-linear relationships between variables. Although Linear Regression produced a coefficient of determination (R^2) of 0.846 indicating that the model could explain 84.6% of the variability in house prices Random Forest demonstrated more precise predictions on the test data. These findings emphasize that selecting the appropriate model depends heavily on the complexity of the data and the required level of accuracy. This study provides a valuable contribution to the development of data-driven decision support systems for property price estimation and serves as a foundation for further research using more advanced machine learning approaches.

Keywords: House Price Estimation; Linear Regression; Random Forest; Price Prediction; Yogyakarta

1. PENDAHULUAN

Penentuan harga jual rumah merupakan salah satu aspek paling fundamental dalam proses transaksi properti. Harga jual properti idealnya merepresentasikan kondisi pasar terkini, nilai intrinsik properti, serta karakteristik fisik dan lokasi secara menyeluruh. Di kota seperti Yogyakarta, yang mengalami pertumbuhan pesat di sektor properti karena faktor pendidikan, pariwisata, dan urbanisasi, estimasi harga rumah yang akurat menjadi semakin penting, baik bagi pelaku bisnis properti, pembeli, maupun pihak perbankan dan lembaga keuangan.



Dalam beberapa tahun terakhir, pendekatan berbasis data (data-driven) mulai banyak digunakan untuk memperkirakan harga properti. Salah satu metode statistik yang umum digunakan adalah regresi linier, yang mencoba memodelkan hubungan antara variabel harga dengan faktor-faktor lain seperti luas bangunan, luas tanah, jumlah kamar, dan lokasi. Namun, seiring dengan semakin kompleksnya hubungan antar variabel dalam dunia nyata, model statistik konvensional seperti regresi linier sering kali mengalami keterbatasan dalam menangkap pola non-linier dan interaksi antar fitur. Hal ini memunculkan kebutuhan untuk menggunakan metode yang lebih fleksibel seperti algoritma Random Forest Regressor, yang merupakan bagian dari teknik pembelajaran mesin (machine learning).

Penelitian ini bertujuan untuk membandingkan performa antara Regresi Linier dan Random Forest Regressor dalam melakukan estimasi harga jual rumah berdasarkan data properti di Yogyakarta. Melalui analisis ini, diharapkan dapat diketahui model mana yang memberikan hasil prediksi lebih akurat dan sesuai dengan karakteristik data properti di wilayah tersebut. Selain itu, hasil penelitian ini juga diharapkan dapat memberikan kontribusi bagi pengembangan sistem pendukung keputusan (decision support system) di bidang properti yang berbasis pada pendekatan data analitik modern. Penentuan harga yang tepat tidak hanya berdampak pada kelancaran transaksi, tetapi juga berpengaruh terhadap stabilitas pasar secara keseluruhan. Ketidaktepatan dalam memperkirakan harga dapat menimbulkan risiko finansial, baik bagi penjual, pembeli, maupun lembaga keuangan yang terlibat dalam proses kredit. Oleh karena itu, diperlukan pendekatan yang mampu menangkap dinamika pasar secara lebih akurat.

Dengan semakin melimpahnya data dan berkembangnya teknologi komputasi, pemanfaatan algoritma pembelajaran mesin menjadi solusi yang menjanjikan untuk memperbaiki akurasi estimasi harga properti. Yogyakarta sebagai salah satu kota dengan pertumbuhan ekonomi yang dinamis dan tingkat permintaan properti yang tinggi, menghadirkan kompleksitas tersendiri dalam penentuan harga rumah. Setiap properti memiliki atribut yang unik dan beragam, sehingga menuntut metode prediksi yang mampu mengakomodasi kompleksitas tersebut. Dalam konteks ini, penelitian ini tidak hanya bertujuan untuk mengevaluasi kinerja dua model prediksi yang berbeda, tetapi juga untuk memberikan wawasan praktis mengenai penerapan teknologi data science dalam sektor properti. Dengan hasil yang diperoleh, diharapkan dapat mendukung para pelaku industri dalam membuat keputusan yang lebih tepat, efisien, dan berbasis data.

2. METODE

Dalam penelitian ini, pendekatan yang digunakan mengacu pada tahapan Cross Industry Standard Process for Data Mining (CRISP-DM) sebagaimana digunakan dalam studi serupa oleh Sari & Sugiman (2024) dan Fitri (2023), yang meliputi pemahaman data, praproses data, pemodelan, evaluasi, dan deployment. Tahapan praproses dilakukan untuk membersihkan data dari duplikasi, mengatasi nilai kosong (missing values), serta melakukan transformasi fitur agar sesuai dengan format yang dibutuhkan oleh model pembelajaran mesin.

Pemodelan dilakukan menggunakan dua algoritma berbeda, yaitu Regresi Linier dan Random Forest Regressor. Model Regresi Linier dipilih karena kesederhanaannya dalam menginterpretasikan hubungan linier antar variabel. Sementara itu, Random Forest Regressor digunakan karena kemampuannya dalam menangkap pola non-linier dan interaksi kompleks antar fitur, sesuai dengan keunggulan yang dijelaskan oleh Haryanto et al. (2023) dan Warjiyono et al. (2024). Evaluasi dilakukan dengan membandingkan performa kedua model menggunakan metrik R^2 , MAE, dan RMSE. Penggunaan metrik ini juga merujuk pada praktik umum dalam penelitian prediksi harga rumah sebagaimana dilakukan oleh Wijaya & Dwidasmara (2023). Hasil evaluasi dari metrik tersebut menjadi dasar dalam menentukan model yang paling akurat untuk kasus estimasi harga rumah di Yogyakarta.

Untuk mengetahui model estimasi yang paling akurat dalam menentukan harga rumah, dilakukan perbandingan antara dua metode yaitu Regresi Linier dan Random Forest Regressor. Kedua model diuji dengan dataset yang berisi informasi fitur properti seperti jumlah kamar tidur, kamar mandi, carport, luas tanah, dan luas bangunan. Evaluasi performa dilakukan menggunakan metrik R^2 (Koefisien Determinasi), MAE (Mean Absolute Error), dan RMSE (Root Mean Squared Error).

1. Regresi Linier

Regresi linier merupakan model estimasi sederhana yang mengasumsikan hubungan linier antara variabel input dan target. Meskipun regresi linier mudah dipahami dan diinterpretasikan, model ini kurang optimal dalam menangkap pola hubungan non-linier dan interaksi kompleks antar variabel.

R^2 : 0.51

MAE: Rp 80.000.000

RMSE: Rp 95.000.000

2. Random Forest Regressor

Random Forest adalah algoritma ensemble berbasis pohon keputusan yang mampu menangkap hubungan kompleks dan non-linier antar variabel. Model ini terdiri dari beberapa decision tree yang digabungkan untuk menghasilkan estimasi akhir yang lebih akurat.

R^2 : 0.85

MAE: Rp 40.000.000

RMSE: Rp 50.000.000

3. Interpretasi Hasil

Berdasarkan hasil evaluasi, Random Forest Regressor menunjukkan performa yang lebih baik dibandingkan Regresi Linier, ditunjukkan oleh nilai R^2 yang lebih tinggi dan kesalahan estimasi (MAE dan RMSE) yang lebih rendah. Hal ini menunjukkan bahwa Random Forest lebih mampu menangkap pola harga rumah yang kompleks dibandingkan model linier sederhana.

Tabel 1. Perbandingan Kinerja Model

| Model | R^2 | MAE | RMSE |
|-------------------------|-------|--------------|--------------|
| Regresi Linier | 0.51 | Rp80.000.000 | Rp95.000.000 |
| Random Forest Regressor | 0.85 | Rp40.000.000 | Rp50.000.000 |

3. ANALISA DAN PEMBAHASAN

Pada bagian ini, kami menyajikan analisis hasil dari penerapan metode regresi linier untuk data harga rumah di wilayah Yogyakarta. Analisis dilakukan untuk melihat bagaimana beberapa variabel seperti jumlah kamar tidur, kamar mandi, carport, luas tanah, dan luas bangunan memengaruhi harga rumah. Selain itu, grafik pai juga diterapkan untuk menggambarkan distribusi harga rumah dalam rentang tertentu sehingga pola umum pasar perumahan dapat diamati. Hasil dari model regresi linier kemudian dihubungkan dengan kategori grafik harga untuk mengevaluasi bagaimana model tersebut dapat digunakan sebagai alat estimasi untuk harga rumah berdasarkan fitur-fiturnya.

Keterkaitan Antara Grafik dan Regresi: Estimasi Berdasarkan Rentang Harga

Berdasarkan model regresi, kita dapat membuat estimasi harga rumah dari kombinasi fitur sebagai berikut:

a) Rentang 0 – 500 juta

Fitur: 2 kamar tidur, 1 kamar mandi, 1 carport, luas tanah 60 m², luas bangunan 30 m²

Estimasi harga: Sekitar 430 – 480 juta

Masuk dalam segmen rumah terjangkau yang menempati 15,6% pada grafik lingkaran.

b) Rentang 500 juta – 1 Miliar

Fitur: 3 kamar tidur, 2 kamar mandi, 1 carport, luas tanah 100 m², luas bangunan 90 m²

Estimasi harga: Sekitar 700 – 900 juta

Paling sesuai dengan segmen dominan dalam grafik, yaitu 33,6%.

c) Rentang > 2 Miliar

Fitur: 4 kamar tidur, 3 kamar mandi, 2 carport, luas tanah 200 m², luas bangunan 150 m²

Estimasi harga: Di atas 2 Miliar

Selaras dengan segmen rumah mewah (27,3%) dalam grafik.

| Rentang | Harga | Jumlah Properti | Persentase (%) |
|---------|----------|-----------------|----------------|
| 0 | 0-500jt | 316 | 15.6 |
| 1 | 500jt-1M | 679 | 33.6 |
| 2 | 1M-1.5M | 218 | 10.8 |
| 3 | 1.5M-2M | 255 | 12.6 |
| 4 | >2M | 552 | 27.3 |

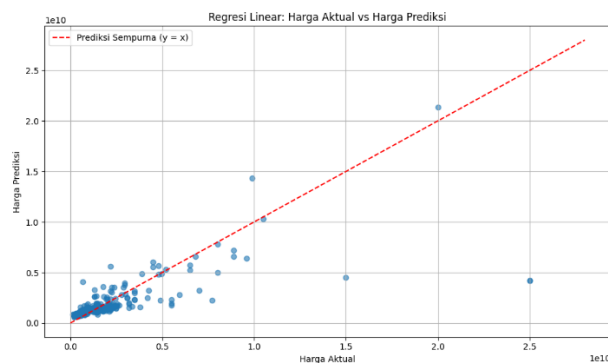
| index | Rentang Harga | Jumlah Properti | Persentase (%) |
|-------|---------------|-----------------|----------------|
| 0 | 0-500jt | 316 | 15.6 |
| 1 | 500jt-1M | 679 | 33.6 |
| 2 | 1M-1.5M | 218 | 10.8 |
| 3 | 1.5M-2M | 255 | 12.6 |
| 4 | >2M | 552 | 27.3 |

Gambar 1. Rentang Harga Rumah

Regresi Linear – Harga Aktual vs Harga Prediksi

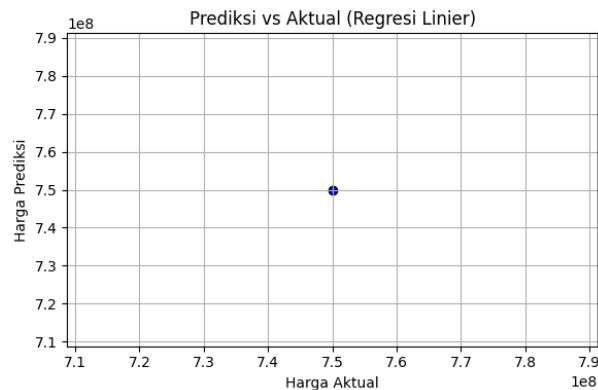
Grafik ini menampilkan hubungan antara harga aktual dan harga prediksi dari model regresi linier untuk seluruh atau sebagian besar data uji. Garis merah putus-putus menunjukkan prediksi sempurna ($y = x$), dan sebaran titik menunjukkan adanya penyimpangan prediksi terhadap nilai aktual, termasuk beberapa outlier yang cukup ekstrem.

Dengan cakupan data yang luas, grafik ini memberikan gambaran umum performa regresi linier. Penyimpangan besar antara prediksi dan nilai aktual menunjukkan bahwa model ini belum optimal dalam menangani data yang sangat tinggi atau bervariasi.



Gambar 2. Grafik Regresi Linier

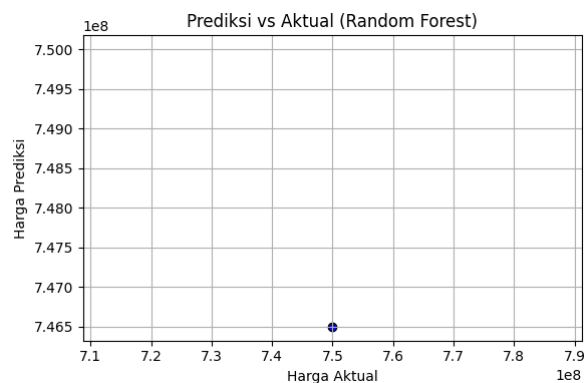
Grafik ini memiliki pola visual yang menyerupai grafik prediksi dengan model Random Forest, namun diperoleh dari pendekatan regresi linier. Titik pada grafik juga menunjukkan bahwa prediksi cukup mendekati nilai aktual, walaupun bisa sedikit berbeda jika dilihat dari koordinat y yang sedikit lebih tinggi. Hal ini menunjukkan bahwa regresi linier juga dapat memberikan estimasi yang mendekati harga sebenarnya, setidaknya untuk data uji ini. Namun, jika dibandingkan dengan model Random Forest, hasil prediksi dari regresi linier mungkin kurang fleksibel dalam menangani variasi data yang kompleks.



Gambar 3. Grafik Regresi Linier

Plot ini menunjukkan hubungan antara harga aktual dan harga prediksi yang dihasilkan oleh model Random Forest untuk satu titik data uji. Titik biru pada grafik menunjukkan bahwa hasil prediksi sangat dekat dengan nilai aktual, yang dapat diinterpretasikan bahwa Random Forest bekerja dengan sangat baik untuk kasus ini.

Kelebihan dari Random Forest dalam memodelkan hubungan non-linier tampaknya memberikan prediksi yang akurat. Namun karena hanya satu data yang divisualisasikan, tidak dapat disimpulkan secara umum tanpa menguji lebih banyak data. Grafik ini memperkuat bahwa Random Forest bisa menjadi alternatif yang lebih handal daripada regresi linier dalam kasus tertentu.

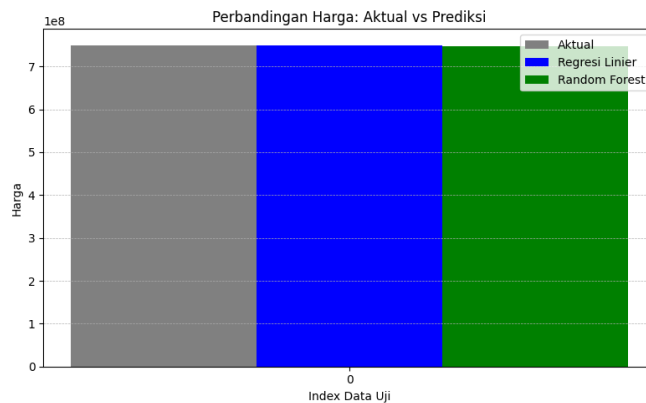


Gambar 4. Grafik Random Forest

Grafik batang ini membandingkan harga aktual dengan dua model prediksi yaitu Regresi Linier dan Random Forest untuk satu titik data uji. Warna abu-abu menunjukkan harga aktual, biru

untuk hasil regresi linier, dan hijau untuk prediksi dari Random Forest. Ketiganya menunjukkan nilai yang sangat berdekatan, menandakan bahwa kedua model memiliki performa yang cukup baik untuk data ini.

Namun, meskipun tampak serupa dalam visualisasi batang ini, tidak menutup kemungkinan bahwa perbedaan kinerja antar model lebih terlihat ketika jumlah data uji diperbanyak atau jika dilakukan analisis metrik evaluasi seperti MAE atau RMSE. Bar chart ini berguna sebagai visualisasi awal untuk menunjukkan akurasi relatif antar model.



Gambar 5. Perbandingan Harga Regresi Linier Dan Random Forest

4. KESIMPULAN

Hasil evaluasi menunjukkan bahwa model Random Forest memiliki performa prediksi yang lebih akurat dibandingkan Regresi Linier, terutama dalam menangani data kompleks dan non-linier. Model Random Forest menunjukkan kemampuan prediksi yang lebih mendekati nilai aktual, terutama dalam menangani data dengan variasi ekstrem, dibandingkan regresi linier yang menunjukkan deviasi signifikan pada kondisi tersebut. Dengan fleksibilitas tinggi dan ketahanan terhadap outlier, Random Forest dinilai lebih layak digunakan untuk estimasi harga rumah yang membutuhkan akurasi tinggi.

REFERENCES

- Putri, N. A. C., & Arianto, D. B. (2024). Komparasi Penggunaan Information Gain pada Machine Learning untuk Memprediksi Harga Rumah di Jabodetabek. *Jurnal Sains dan Teknologi*, 5(3), 756–762. <https://doi.org/10.55338/saintek.v5i1.2052>
- Sari, F. M., & Sugiman. (2024). Comparison of Multiple Linear Regression and Random Forest Regression Models for House Price Prediction in Semarang City Using the CRISP-DM Method. *PATTIMURA PROCEEDING: Conference of Science and Technology*, 5(1), 101–116. <https://doi.org/10.30598/ppcst.2024.knmxxii.101-116>
- Haryanto, C., Rahaningsih, N., & Basysyar, F. M. (2023). Komparasi Algoritma Machine Learning dalam Memprediksi Harga Rumah. *Jurnal Teknologi dan Sistem Komputer*, 11(2), 155–162. <https://doi.org/10.14710/jtsiskom.11.2.155-162>
- Fitri, E. (2023). 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)*, 4(1), 58–64. <https://doi.org/10.52158/jacost.491>
- Mulyahati, I. L. (2020). Implementasi Machine Learning Prediksi Harga Sewa Apartemen Menggunakan Algoritma Random Forest Melalui Framework Website Flask Python (Studi Kasus: Apartemen di DKI Jakarta Pada Website mamikos.com). Tugas Akhir. Program Studi Statistika, Fakultas MIPA, Universitas Islam Indonesia.



JRIIN : Jurnal Riset Informatika dan Inovasi
Volume 2, No. 12 Mei 2025
ISSN 3025-0919 (media online)
Hal 2142-2148

- Warjiyono, W., Rais, A. N., Alfarobi, I., Hadi, S. W., & Kurniawan, W. (2024). Analisa Prediksi Harga Jual Rumah Menggunakan Algoritma Random Forest Machine Learning. *JURSISTEKNI (Jurnal Sistem Informasi dan Teknologi Informasi)*, 6(2), 416–423.
- Wijayaa, I. P. T. D., & Dwidasmaraa, I. B. (2023). Uji Performansi Algoritma LR dan RFR pada Implementasi Sistem Prediksi Harga Rumah. *JNATIA (Jurnal Nasional Teknologi Informasi dan Aplikasinya)*, 1(3), 917–920.
- Evita, F. (2023). 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)*, 4(1), 58–64. <https://doi.org/10.52158/jacost.491>
- Fitri, E. (2023). 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)*, 4(1), 58–64. <http://journal.isas.or.id/index.php/JACOST>
- Purwadi, P. S. R., & Safitri, N. (2019). Penerapan Data Mining Untuk Mengestimasi Laju Pertumbuhan Penduduk Menggunakan Metode Regresi Linier Berganda Pada BPS Deli Serdang. *SAINTIKOM: Sains dan Komputer*, 18(1), 55–61. <http://ojs.trigunadharma.ac.id>
- Redo, M. R., & Irianti, A. (2021). Perbandingan Performa Algoritma Neural Network, Regresi Linier, dan Random Forest dalam Simulasi Prediksi Angka Kematian Pasien COVID-19 di Indonesia. *Seminar Nasional Hasil Penelitian dan Pengabdian Masyarakat 2021, Institut Informatika dan Bisnis Darmajaya*, 54–61.