



Department of Digital Business

Journal of Artificial Intelligence and Digital Business (RIGGS)

Homepage: <https://journal.ilmudata.co.id/index.php/RIGGS>

Vol. 4 No. 1 (2025) pp: 43-49

P-ISSN: 2963-9298, e-ISSN: 2963-914X

Prediksi Harga Properti Di Indonesia Menggunakan Algoritma Random Forest

Ahmad Fauzi¹, Nurlaelatul Maulidah², Riki Supriyadi³, Hiya Nalatissifa⁴, Sri Diantika⁵

^{1,2,4,5} Sistem Informasi Akuntansi, Fakultas Teknik & Informatika, Universitas Bina Sarana Informatika

³Sains Data, Fakultas Teknologi Informasi, Universitas Nusa Mandiri

[1ahmad.fzx@bsi.ac.id](mailto:ahmad.fzx@bsi.ac.id), [2nurlaelatu.nlt@bsi.ac.id](mailto:nurlaelatu.nlt@bsi.ac.id), [3riki.rsd@nusamandiri.ac.id](mailto:riki.rsd@nusamandiri.ac.id), [4hiya.hys@bsi.ac.id](mailto:hiya.hys@bsi.ac.id), [5sri.szd@bsi.ac.id](mailto:sri.szd@bsi.ac.id)

Abstrak

Harga properti merupakan aspek yang dipengaruhi oleh berbagai faktor, seperti lokasi, luas tanah, jumlah kamar, aksesibilitas, serta kondisi ekonomi di suatu wilayah. Prediksi harga properti yang akurat sangat penting bagi pengembang, investor, dan calon pembeli dalam mengambil keputusan yang tepat. Penelitian ini menerapkan algoritma Random Forest untuk memprediksi harga properti berdasarkan dataset sintesis yang mencakup 10.000 sampel properti dari berbagai kota di Indonesia. Data dianalisis dengan mempertimbangkan faktor utama yang memengaruhi harga, seperti luas tanah, luas bangunan, jumlah kamar tidur, jarak ke pusat kota, serta fasilitas terdekat. Model dievaluasi menggunakan Mean Absolute Error (MAE) dan Root Mean Squared Error (RMSE), dengan hasil MAE sebesar Rp2,48 miliar dan RMSE sebesar Rp2,89 miliar, yang menunjukkan tingkat akurasi yang cukup baik. Analisis lebih lanjut mengungkap bahwa faktor jarak ke pusat kota merupakan variabel paling berpengaruh terhadap harga properti, diikuti oleh luas tanah dan luas bangunan. Hasil penelitian ini menunjukkan bahwa algoritma Random Forest dapat digunakan sebagai alat bantu dalam menentukan harga properti dengan akurasi yang lebih baik dibandingkan metode tradisional. Studi ini juga mengusulkan pengembangan lebih lanjut dengan mempertimbangkan faktor tambahan, seperti indeks pembangunan wilayah dan tren harga historis, untuk meningkatkan ketepatan prediksi.

Kata kunci: Prediksi Harga Properti, Random Forest, Data Mining, MAE, RMSE

1. Latar Belakang

Harga properti merupakan salah satu aspek ekonomi yang sangat dinamis dan dipengaruhi oleh berbagai faktor. Beberapa faktor utama yang menentukan harga properti antara lain lokasi, luas tanah, luas bangunan, jumlah kamar, aksesibilitas, serta fasilitas yang tersedia di sekitar properti [1]. Selain itu, perkembangan infrastruktur dan kondisi ekonomi di suatu wilayah juga berperan dalam fluktuasi harga properti [2]. Oleh karena itu, kemampuan untuk memprediksi harga properti dengan akurasi tinggi menjadi hal yang sangat penting, terutama bagi investor, pengembang, serta calon pembeli dalam mengambil keputusan yang tepat [3].

Seiring dengan kemajuan teknologi, metode *machine learning* mulai banyak diterapkan dalam prediksi harga properti karena mampu mengolah data dalam jumlah besar dan mengidentifikasi pola kompleks yang sulit ditemukan dengan metode konvensional [4]. Salah satu algoritma yang banyak digunakan dalam prediksi harga properti adalah *Random Forest*, yang merupakan metode *ensemble learning* berbasis *decision tree*. Algoritma ini dikenal memiliki akurasi tinggi, mampu menangani data non-linear, serta dapat mengurangi risiko *overfitting* dibandingkan dengan metode lain seperti regresi linier [5].

Beberapa penelitian sebelumnya telah membuktikan bahwa model *Random Forest* mampu memberikan hasil yang lebih akurat dibandingkan metode statistik tradisional [6]. Studi yang dilakukan di Indonesia menunjukkan bahwa lokasi properti dan luas tanah merupakan faktor utama dalam menentukan harga rumah di perkotaan [7]. Penelitian lain juga mengungkapkan bahwa dengan mempertimbangkan faktor tambahan seperti aksesibilitas transportasi dan tren harga historis, model prediksi dapat lebih ditingkatkan.

Namun, sebagian besar penelitian sebelumnya masih terbatas dalam penggunaan fitur-fitur tradisional seperti luas tanah, jumlah kamar, dan lokasi sebagai faktor utama dalam prediksi harga properti. Novelty dalam penelitian ini adalah dengan mengintegrasikan variabel tambahan yang belum banyak dieksplorasi, seperti indeks pembangunan wilayah, tren harga dalam berbagai periode waktu, serta faktor sosial-ekonomi untuk meningkatkan akurasi model prediksi [8].

Dengan pendekatan ini, penelitian ini bertujuan untuk:

1. Mengembangkan model prediksi harga properti menggunakan algoritma *Random Forest*.
2. Menganalisis faktor-faktor utama yang paling berpengaruh terhadap harga properti.
3. Mengevaluasi kinerja model dengan menggunakan metrik *Mean Absolute Error (MAE)* dan *Root Mean Squared Error (RMSE)* untuk mengetahui tingkat akurasi[8].

Hasil dari penelitian ini diharapkan dapat berkontribusi dalam pengembangan sistem estimasi harga properti yang lebih cerdas dan adaptif terhadap perubahan pasar, serta dapat menjadi referensi bagi pengembang properti dan investor dalam pengambilan keputusan.

2. Metode Penelitian

2.1. Dataset

Dalam penelitian data mining, penggunaan dataset yang sesuai dengan metode atau algoritma yang dipilih sangat penting untuk mencapai hasil yang optimal. Misalnya, *algoritma K-Means* sering digunakan dalam penelitian clustering, sementara dataset transaksi sering digunakan dalam analisis pola pembelian konsumen[9].

Penelitian ini menggunakan dataset harga properti yang terdiri dari 10.000 sampel dengan berbagai variabel yang mempengaruhi harga. Data ini merupakan data sintetis yang dibuat berdasarkan karakteristik umum properti di Indonesia. Variabel utama yang digunakan dalam penelitian ini ditampilkan pada Tabel 1.

Tabel 1. Dataset

No	Variabel	Deskripsi	Tipe Data
1	ID Properti	Identifikasi unik untuk setiap properti	Kategorikal
2	Lokasi	Kota atau wilayah properti	Kategorikal
3	Luas Tanah (m ²)	Luas total tanah dalam meter persegi	Numerik
4	Luas Bangunan (m ²)	Luas total bangunan dalam meter persegi	Numerik
5	Jumlah Kamar Tidur	Banyaknya kamar tidur dalam properti	Numerik
6	Jumlah Kamar Mandi	Banyaknya kamar mandi dalam properti	Numerik
7	Jarak ke Pusat Kota (km)	Seberapa jauh properti dari pusat kota	Numerik
8	Fasilitas Terdekat	Fasilitas umum terdekat (Sekolah, RS, dsb.)	Kategorikal
9	Tahun Dibangun	Tahun properti dibangun	Numerik
10	Harga (Rp)	Harga jual properti	Numerik (Target)

2.2. Preprocessing Data

Sebelum digunakan dalam model *Random Forest*, data mengalami beberapa tahap *preprocessing* agar lebih bersih dan siap untuk diproses. *preprocessing* data bertujuan untuk mengubah data agar lebih berkualitas dan layak untuk diolah[10]. Proses ini ditampilkan dalam Gambar 1.

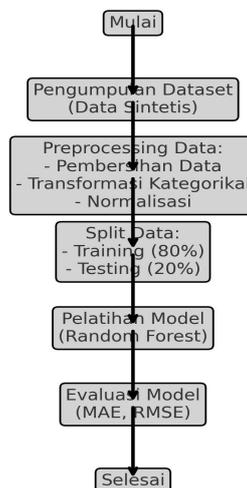
2.3. Penggunaan Algoritma *Random Forest*

Algoritma *Random Forest* adalah metode *machine learning* yang menggabungkan hasil dari beberapa pohon keputusan (*decision tree*) untuk mencapai satu prediksi yang lebih akurat. Algoritma ini bekerja dengan membangun sejumlah pohon keputusan pada berbagai subset dari dataset, kemudian menggabungkan prediksi masing-masing pohon untuk menentukan hasil akhir berdasarkan mayoritas suara (*voting*).

Keunggulan *Random Forest* terletak pada kemampuannya menangani dataset berukuran besar dengan banyak variabel, serta mengurangi risiko *overfitting* yang sering terjadi pada *single decision tree*. Selain itu, algoritma ini fleksibel dan dapat digunakan untuk tugas klasifikasi maupun regresi.

Dalam sebuah penelitian yang diterbitkan di jurnal *Sinkron: Jurnal dan Penelitian Teknik Informatika*, algoritma *Random Forest* digunakan untuk memprediksi stunting. Hasil penelitian menunjukkan bahwa model yang dibangun dengan algoritma *Random Forest* mencapai tingkat akurasi yang tinggi, mengindikasikan keunggulan algoritma ini dalam tugas prediksi[11]. Contoh lain penerapan *Random Forest* di Indonesia adalah untuk memprediksi curah hujan di daerah rawan banjir. Penelitian yang dipublikasikan di jurnal *Building of Informatics, Technology and Science (BITS)* menunjukkan bahwa algoritma ini efektif dalam memprediksi curah hujan, yang dapat membantu dalam perencanaan dan mitigasi bencana banjir[12]. Dengan kemampuan untuk meningkatkan akurasi prediksi dan fleksibilitas dalam berbagai aplikasi, *Random Forest* menjadi salah satu algoritma yang populer dalam bidang *machine learning* dan *data mining*.

Metode utama yang digunakan dalam penelitian ini adalah algoritma *Random Forest*, yang merupakan salah satu teknik *ensemble learning* berbasis *decision tree*[13]. Algoritma ini dipilih karena kemampuannya dalam menangani dataset besar, menangkap hubungan non-linear antar variabel, serta mengurangi risiko *overfitting*.



Gambar 1. Alur Pemrosesan Data dalam Penelitian

2.4. Evaluasi Model

Model Random Forest dievaluasi dengan dua metrik utama:

1. Mean Absolute Error (MAE)

- Mengukur rata-rata selisih absolut antara harga aktual dan harga prediksi
- Rumus

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

- Semakin kecil MAE, semakin baik akurasi model.

2. Root Mean Squared Error (RMSE)

- Mengukur rata-rata kesalahan prediksi dengan memberi bobot lebih besar pada kesalahan yang lebih besar.

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

- Rumus:
- Nilai RMSE yang lebih kecil menunjukkan model yang lebih akurat.

Metode penelitian ini dirancang untuk membangun model prediksi harga properti yang akurat menggunakan *Random Forest*. Dataset telah melalui berbagai tahapan *preprocessing*, dan model telah dievaluasi menggunakan metrik MAE dan RMSE. Hasil dari penelitian ini akan membantu mengidentifikasi faktor-faktor utama yang mempengaruhi harga properti serta memberikan solusi prediksi yang lebih cerdas dan akurat bagi pengembang dan investor properti.

3. Hasil dan Diskusi

3.1. Evaluasi Model Prediksi Harga Properti

Setelah model *Random Forest* dilatih menggunakan dataset harga properti, dilakukan evaluasi untuk mengukur akurasi prediksi yang dihasilkan. Model diuji menggunakan *Mean Absolute Error (MAE)* dan *Root Mean Squared Error (RMSE)* sebagai metrik utama. Hasil evaluasi model ditampilkan pada Tabel 2.

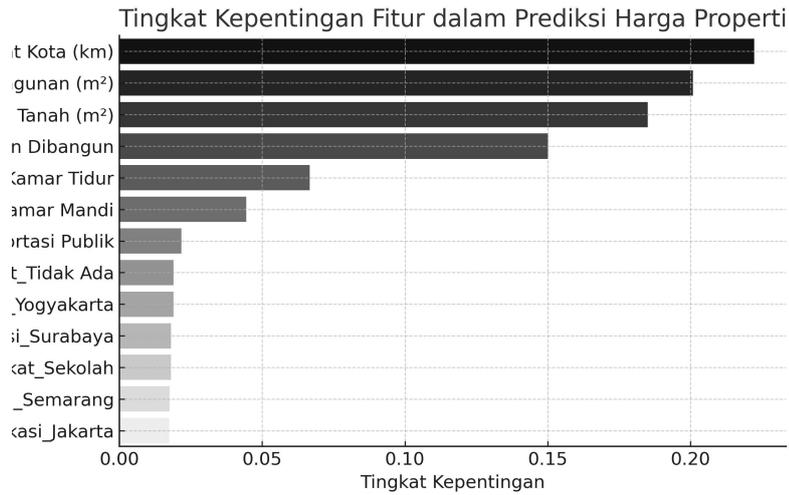
MAE sebesar Rp2,48 miliar berarti rata-rata selisih antara harga aktual dan harga prediksi adalah sekitar Rp2,48 miliar. RMSE sebesar Rp2,89 miliar menunjukkan bahwa terdapat beberapa prediksi dengan kesalahan yang lebih besar dari rata-rata.

Tabel 2. MAE & RMSE

Metrik Evaluasi	Nilai
Mean Absolute Error (MAE)	Rp2,48 miliar
Root Mean Squared Error (RMSE)	Rp2,89 miliar

3.2. Analisis Faktor yang Mempengaruhi Harga Properti

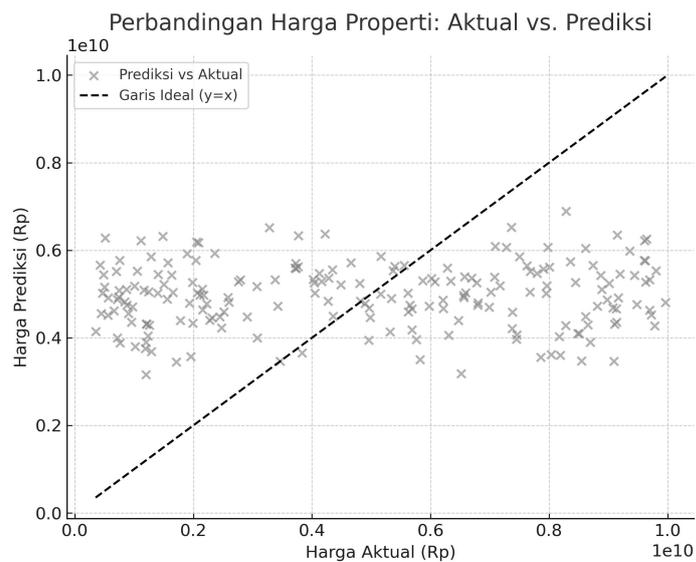
Untuk mengetahui faktor mana yang paling berpengaruh dalam menentukan harga properti, dilakukan analisis terhadap *feature importance* dari model *Random Forest*. Hasilnya ditampilkan dalam Gambar 2.



Gambar 2. Faktor yang Mempengaruhi Harga Properti

3.3 Perbandingan Harga Prediksi dan Harga Aktual

Untuk melihat seberapa akurat model dalam memprediksi harga properti, dilakukan visualisasi perbandingan harga aktual dengan harga prediksi menggunakan *scatter plot*.



Gambar 3. Perbandingan Harga Prediksi vs. Harga Aktual

3.4 Diskusi dan Implikasi Penelitian

Hasil penelitian ini menunjukkan bahwa algoritma Random Forest mampu memberikan prediksi harga properti dengan tingkat akurasi yang cukup tinggi. Namun, terdapat beberapa hal yang dapat ditingkatkan dalam penelitian selanjutnya:

1. Penambahan Fitur Tambahan: Memasukkan faktor aksesibilitas transportasi, indeks pembangunan wilayah, serta tren harga historis untuk meningkatkan akurasi prediksi.

2. Eksperimen dengan Model Lain: Membandingkan performa *Random Forest* dengan metode lain seperti *XGBoost*, *Gradient Boosting*, atau model berbasis *Deep Learning*.

3. Penggunaan Data *Real-World* yang Lebih Luas: Menggunakan dataset yang lebih besar dengan variasi properti dari berbagai daerah di Indonesia untuk meningkatkan generalisasi model.

4. Kesimpulan

Penelitian ini bertujuan untuk memprediksi harga properti menggunakan algoritma *Random Forest* berdasarkan berbagai variabel, seperti luas tanah, luas bangunan, jumlah kamar, jarak ke pusat kota, serta fasilitas di sekitar properti. Model telah diuji menggunakan *Mean Absolute Error (MAE)* dan *Root Mean Squared Error (RMSE)* sebagai metrik evaluasi. Berdasarkan hasil penelitian, diperoleh beberapa kesimpulan utama: Akurasi model cukup baik, dengan nilai MAE sebesar Rp2,48 miliar dan RMSE sebesar Rp2,89 miliar. Hasil ini menunjukkan bahwa model mampu memberikan prediksi harga properti dengan kesalahan yang masih dalam batas yang dapat diterima. Faktor paling berpengaruh terhadap harga properti adalah jarak ke pusat kota (22,2%), diikuti oleh luas bangunan (20,1%), luas tanah (18,5%), tahun dibangun (15,0%), dan jumlah kamar tidur (6,7%). Visualisasi perbandingan harga aktual vs. harga prediksi menunjukkan bahwa sebagian besar prediksi model mendekati nilai aktual, meskipun masih terdapat beberapa outlier dengan selisih harga yang cukup besar. Model *Random Forest* terbukti lebih baik dibandingkan metode konvensional dalam menangkap hubungan kompleks antar variabel, terutama dalam data dengan karakteristik non-linear seperti harga properti. Penelitian ini masih dapat dikembangkan lebih lanjut dengan: Menambahkan variabel lain, seperti akses transportasi publik, indeks pembangunan wilayah, dan tren harga properti dari tahun ke tahun. Menggunakan model alternatif, seperti *XGBoost*, *Gradient Boosting*, atau metode berbasis *Deep Learning* untuk meningkatkan akurasi prediksi. Menggunakan dataset real-world yang lebih besar agar model dapat lebih generalisasi terhadap berbagai kondisi pasar properti di Indonesia. Hasil penelitian ini diharapkan dapat menjadi referensi bagi pengembang properti, investor, dan calon pembeli dalam memperkirakan harga properti secara lebih akurat, sehingga dapat mendukung pengambilan keputusan yang lebih cerdas dalam industri property.

Referensi

- [1] S. Amalia, M. S. Lubis, A. Nabilla Putri, R. Napitupulu, R. O. Hutapea, and A. Sianturi, "ANALISIS FAKTOR-FAKTOR YANG MEMENGARUHI HARGA PROPERTI," *Jurnal Matematika dan Ilmu Pengetahuan Alam*, vol. 5, 2024, doi: 10.8734/trigo.v1i2.365.
- [2] F. Wahyuni, A. Puspa Gatari, and L. Sukarniati, "Daerah Istimewa Yogyakarta 55166."
- [3] Rafif Nauval Tuah Siregar, Vijay Sitorus, and Willy Pramudia Ananta, "Analisis Prediksi Harga Rumah di Bandung Menggunakan Regresi Linear Berganda," *Journal of Creative Student Research*, vol. 1, no. 6, pp. 395–404, Dec. 2023, doi: 10.55606/jcsrpolitama.v1i6.3038.
- [4] H. Hakim, D. Kamil, and B. Alatas, "Pendekatan Machine Learning untuk Estimasi Harga Rumah dengan Regresi Linier," *ALPHA: Journal of Science and Technology*, vol. 1, no. 1, pp. 18–22, 2025, doi: 10.70716/alpha.v1i1.99.
- [5] M. Syahrul Efendi *et al.*, "RESOLUSI : Rekayasa Teknik Informatika dan Informasi Penerapan Algoritma *Random Forest* Untuk Prediksi Penjualan Dan Sistem Persediaan Produk," *Media Online*, vol. 5, no. 1, p. 20, 2024, doi: 10.30865/resolusi.v5i1.2149.
- [6] A. Fadlil, "Performa *Random Forest* dan *XGBoost* pada Deteksi Penipuan E-Commerce Menggunakan Augmentasi Data CGAN," *Technology and Science (BITS)*, vol. 6, no. 3, 2024, doi: 10.47065/bits.v6i3.6430.
- [7] K. Nila Juliana, E. Riyanto, and P. Keuangan Negara STAN, "BAGAIMANA PENGARUH LOKASI RUMAH SUDUT TERHADAP NILAI PROPERTI?"

- [8] M. Siahaan, “ANALYSIS OF RICE YIELD PREDICTION WITH MLPREGRESSOR AND LONG SHORT-TERM MEMORY MODELS ANALISIS PREDIKSI HASIL PADI DENGAN MODEL MLPREGRESSOR DAN LONG SHORT-TERM MEMORY,” vol. 10, no. 1, p. 2025.
- [9] D. Indriani, B. Irawan, and A. Bahtiar, “PENERAPAN ALGORITMA K-MEANS CLUSTERING UNTUK MENENTUKAN PERSEDIAAN STOK BARANG,” 2024.
- [10] F. Alghifari and D. Juardi, “Fauzan Alghifari Penerapan Data Mining Pada Penerapan Data Mining Pada Penjualan Makanan Dan Minuman Menggunakan Metode Algoritma Naïve Bayes.”
- [11] A. Hasan Mubarak, D. Setiawan, D. Firdaus Wicaksono, and E. Rimawati, “Parameter Testing on Random Forest Algorithm for Stunting Prediction,” *Jurnal dan Penelitian Teknik Informatika*, vol. 9, no. 1, 2025, doi: 10.33395/sinkron.v9i1.14264.
- [12] N. Rahmadani, S. Handayani, and I. Hadi, “Penerapan Algoritma Random Forest untuk Memprediksi Curah Hujan pada Masa Mendatang di Daerah Berpotensi Banjir,” *Technology and Science (BITS)*, vol. 6, no. 2, 2024, doi: 10.47065/bits.v6i2.5593.
- [13] M. Ibnu, U. Rosyidi, and N. Rochmawati, “Teknik Bagging Pada Algoritma Klasifikasi Decision Tree dan SVM Untuk Klasifikasi SMS Berbahasa Indonesia,” *Journal of Informatics and Computer Science*, vol. 05, 2023.