

Implementasi *Machine Learning* untuk Memprediksi Harga Rumah dengan Menggunakan Metode *Decision Tree Regressor* dan *Random Forest Regressor*

Mawarita Salsabila^{1*}, Rangga Prasetya Adiwijaya², Lia Nur Octavia¹

¹Program Studi Teknik Industri, Fakultas Teknologi dan Bisnis Energi, Institut Teknologi PLN, Jakarta Barat, Jakarta.

²Program Studi Teknik Elektro, Fakultas Ketenagalistrikan dan Energi Terbarukan, Institut Teknologi PLN, Jakarta Barat, Jakarta.

Informasi Artikel

Naskah Diterima :

24 Januari 2025

Direvisi : 25 Mei 2025

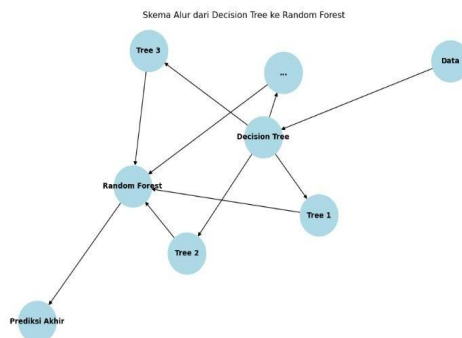
Disetujui : 28 Mei 2025

doi: 10.62870/setrum.v14i1.31025

*Korespondensi Penulis :

mawaritasl21@gmail.com

Graphical abstract



Abstract

Along with the rapid development of technology, the need for efficient digital solutions in various fields is increasing, including in the property sector. Searching and assessing house prices usually requires direct interaction with property agents or homeowners. This creates the need for a more efficient house price prediction method without having to interact directly with the person concerned. This study proposes the implementation of Random Forest Regressor in the Machine Learning framework which is realized in the form of an application to predict house prices based on property images that can be used by anyone and anywhere. The dataset used in this study includes thousands of numerical data such as the number of bedrooms, the number of bathrooms, the area of the building, the condition of the building, and accessibility. The Random Forest Regression method can be applied after the user adds Feature Description and trains the Data Train in Machine Learning to ensure the efficiency and accuracy of this method. The results of the study show that the image-based model with Random Forest Regression with the help of Decision Tree can predict house prices with fairly good accuracy. This application can be accessed by users from various regions, thus providing flexibility and ease of use. This method has proven to be effective for initial analysis and can be a significant tool in making property valuation decisions.

Keywords: House price prediction, Random Forest Regression, Machine Learning, Feature Description, Decision Tree

Abstrak

Seiring dengan perkembangan teknologi yang pesat, kebutuhan akan solusi digital yang efisien dalam berbagai bidang semakin meningkat, termasuk dalam sektor properti. Pencarian dan penilaian harga rumah biasanya memerlukan interaksi langsung dengan agen properti atau pemilik rumah. Hal ini menciptakan kebutuhan akan metode prediksi harga rumah yang lebih efisien tanpa harus berinteraksi langsung dengan yang bersangkutan. Penelitian ini mengusulkan implementasi *Random Forest Regressor* dalam kerangka *Machine Learning* yang diwujudkan dalam bentuk aplikasi untuk memprediksi harga rumah berdasarkan citra properti yang dapat digunakan oleh siapa saja dan di mana saja. Dataset yang digunakan dalam penelitian ini mencakup ribuan data numerik seperti jumlah kamar tidur, jumlah kamar mandi, luas bangunan, kondisi bangunan, dan aksesibilitas. Metode *Random Forest Regression* dapat diterapkan setelah pengguna menambahkan *Feature Description* dan melatih Data Train dalam *Machine Learning* untuk memastikan efisiensi dan akurasi metode ini. Hasil penelitian menunjukkan bahwa model berbasis citra dengan *Random Forest Regression* dengan bantuan *Decision Tree* dapat memprediksi harga rumah dengan akurasi yang cukup baik. Aplikasi ini dapat diakses oleh pengguna dari berbagai wilayah, sehingga memberikan fleksibilitas dan kemudahan dalam

penggunaan. Metode ini terbukti efektif untuk analisis awal dan dapat menjadi alat bantu yang signifikan dalam pengambilan keputusan penilaian properti.

Kata kunci: Prediksi harga rumah, *Random Forest Regression*, *Machine Learning*, *Feature Description*, *Decision Tree*

© 2025 Penerbit Jurusan Teknik Elektro UNTIRTA Press. All rights reserved.

1. PENDAHULUAN

Pertumbuhan penduduk yang semakin pesat akibat imigran atau perantau yang singgah dan tinggal di suatu kota mengakibatkan masyarakat harus memiliki tempat tinggal yang layak yang sesuai dengan anggaran dan kebutuhan rumah tangganya. Harga setiap rumah meningkat setiap tahunnya. Kenaikan harga ini dapat diukur dari berbagai aspek dan fitur yang dimiliki oleh rumah tersebut. Harga dapat selalu berubah sewaktu-waktu dan tidak dapat diprediksi khususnya bagi masyarakat baru yang berpindah kota tempat tinggal dan belum mengetahui harga wajar untuk rumah di wilayah tersebut.

Seiring dengan perkembangan teknologi, khususnya dalam tata cara penjualan properti, biasanya pemilik rumah atau agen properti akan mencantumkan nomor telepon yang dapat dihubungi sebagai jembatan untuk bernegosiasi dalam jual beli. Dengan adanya metode prediksi ini akan membantu dalam pertimbangan untuk pengambilan keputusan berdasarkan fitur properti yang dijual[1]. Penerapan ini cocok digunakan pada pembeli yang sedang mencari rumah sekaligus ingin memperkirakan harganya hanya dengan mendeskripsikan fitur rumah yang ada seperti jumlah kamar, jumlah kamar mandi, luas bangunan, kondisi bangunan dan aksesibilitas tanpa berhubungan langsung dengan agen properti atau pemilik rumah yang akan memakan banyak waktu[2].

Cabang ilmu kecerdasan buatan yang digunakan disini adalah *machine learning*. *Machine Learning* adalah ilmu yang mempelajari tentang bagaimana komputer mampu mempelajari data-data yang diinput untuk diproses ke tahap selanjutnya sebagai *ouput*[3]. *Output* inilah yang akan menunjukkan peningkatan kecerdasannya dibandingkan sebelum adanya *traning* data[4]. Pada penelitian sebelumnya, terdapat penelitian dengan penggunaan *linear regression* tanpa adanya penambahan metode *decision tree* sebelum data memproses deskripsi tentang fitur rumah sehingga akurasi yang didapatkan kurang relevan.

Berdasarkan beberapa penelitian yang telah dilakukan sebelumnya, tujuan penelitian kali ini akan membuat model prediksi harga rumah menggunakan metode *machine learning* dan algoritmanya adalah *Random Forest Regression* dengan adanya elemen-elemen harga rumah rata-rata berbanding dengan fitur rumah sebelum hasil prediksi harga muncul. Manfaat penelitian ini adalah untuk menyelidiki dan menentukan keakuratan penerapan pembelajaran mesin menggunakan algoritma *Random Forest Regreesion* yang akan digunakan untuk memprediksi harga properti jangka panjang.

1.1 Tahapan Data

- 1) Melakukan Pelatihan *Train Data* dengan menambahkan ribuan data numerik yang berasal dari *open source* seperti luas tanah, ukuran rumah, dan akses yang dapat mempengaruhi harga rumah.
- 2) Melakukan klasifikasi perbandingan harga dengan bantuan *Decision Tree Regression* yang dapat membantu metode *Random Forest Regreesion* dalam memprediksi harga.
- 3) Melakukan pengumpulan voting mayoritas dari hasil prediksi semua pohon lalu menghitung rata-rata dari hasil prediksi semua pohon keputusan untuk hasil akhir dari perbandingan harga yang sesuai.

1.2 Machine Learning

Machine Learning (ML) adalah teknik kecerdasan buatan yang memungkinkan sistem untuk belajar dari data dan pengalaman untuk membuat model prediktif dan pengambilan keputusan tanpa deprogram secara eksplisit[5]. *Machine Learning* memungkinkan komputer untuk memprediksi harga rumah dan gambar dengan cara *machine learning* tersebut dilatih untuk mengenali pola visual terkait harga, misalnya elemen arsitektur atau kualitas dari bangunan itu sendiri, dan dapat mengambil keputusan sendiri tanpa campur tangan manusia[6].

1.3 Decision Tree Regression

Decision Tree adalah algoritma pembelajaran mesin yang digunakan dalam penugasan klasifikasi dan regresi, di mana data dipecah menjadi subset berdasarkan fitur-fitur yang dipilih, hingga mencapai node daun yang mewakili keputusan akhir. Struktur ini menyerupai diagram alur, dengan setiap node internal sebagai pengambil keputusan berdasarkan suatu atribut, dan cabang-cabang yang menunjukkan hasil dari keputusan tersebut[7]. Meskipun model *decision tree* sederhana dan mudah dipahami, model ini rentan terhadap *overfitting*, terutama jika diterapkan pada dataset yang besar dan kompleks. Untuk mengatasi keterbatasan ini, *Random Forest Regression* digunakan sebagai metode kombinasi yang menggabungkan banyak *decision trees* independen yang dilatih pada subset data acak[8]. Proses ini meningkatkan akurasi prediksi dan mengurangi risiko *overfitting* dengan mengambil rata-rata hasil dari semua, sehingga menciptakan model yang lebih stabil dan akurat[9]. Sebagai contoh, penelitian oleh Leo Breiman tahun 2001, menekankan bahwa penggabungan model-model sederhana seperti *decision tree* dalam pendekatan gabungan seperti *Random Forest* dapat mengatasi variabilitas data yang tinggi dan meningkatkan kinerja dalam prediksi[10].

1.4 Random Forest Regression

Random Forest Regression adalah sebuah metode ensemble learning yang dikembangkan oleh Leo Breiman pada tahun 2001 untuk tugas regresi. Metode ini menggabungkan hasil dari banyak pohon keputusan (*decision trees*) yang dilatih pada subset data acak, kemudian memberikan *output* berupa rata-rata prediksi. Pendekatan ini memiliki keunggulan dalam mengurangi *overfitting* dan meningkatkan akurasi prediksi, karena prediksi dikombinasikan dari banyak pohon yang ini adalah independen[11]. Metode ini juga dikenal mampu menangani data berukuran besar serta memberikan wawasan tentang fitur yang paling berpengaruh dalam model (*feature importance*).

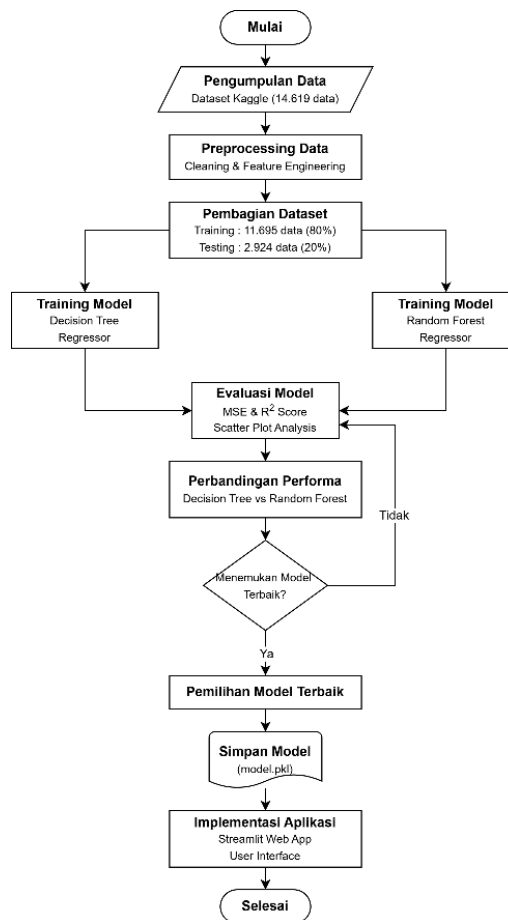
Penelitian oleh Leo Breiman dalam *Random Forests* dan implementasi oleh Liam dan Wiener menunjukkan efektivitas metode ini dalam berbagai aplikasi. Selain itu, analisis oleh Fernandez Delgado tahun 2014, menegaskan bahwa *Random Forest* adalah salah satu algoritma yang paling unggul untuk tugas klasifikasi dan regresi dibandingkan ratusan metode lainnya[12], [13]. Hal ini menjadikan *Random Forest Regression* sebagai salah satu metode andalan dalam memproses data yang kompleks di berbagai bidang.

1.5 Variable Dataset

- 1) Kamar tidur – Jumlah kamar tidur (1-10 atau lebih)
- 2) Kamar mandi – Jumlah kamar mandi (1-10 atau lebih)
- 3) Bangunan – Luas Bangunan Keseluruhan (dihitung dalam M^2)
- 4) Kondisi – Kondisi Bangunan Keseluruhan (diurutkan dari 1-5)
- 5) Aksesibilitas – Jumlah Sekolah Terdekat (1-10 atau lebih)

2. METODE

Dataset yang digunakan dalam penelitian ini diperoleh dari *platform* Kaggle dengan total 14.619 data properti yang mencakup berbagai fitur rumah seperti jumlah kamar tidur, kamar mandi, luas bangunan, kondisi bangunan, dan aksesibilitas sekolah terdekat. Dataset ini dipilih karena memiliki kelengkapan data yang memadai dan representatif untuk analisis prediksi harga rumah.



Gambar 1. Diagram Alur Penelitian

Dalam proses pemodelan, dataset dibagi menjadi dua bagian menggunakan metode *train_test_split* dengan rasio 80:20, yaitu:

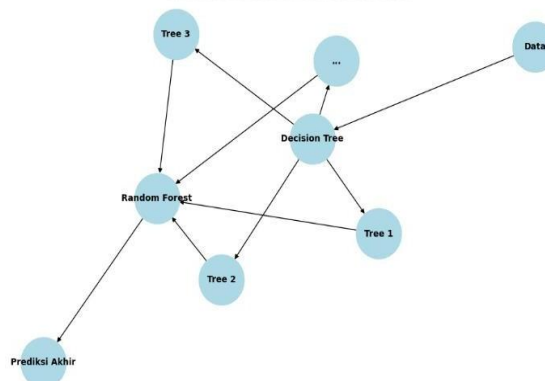
- Data *Training*: 11.695 data (80%) digunakan untuk melatih model
- Data *Testing*: 2.924 data (20%) digunakan untuk menguji performa model

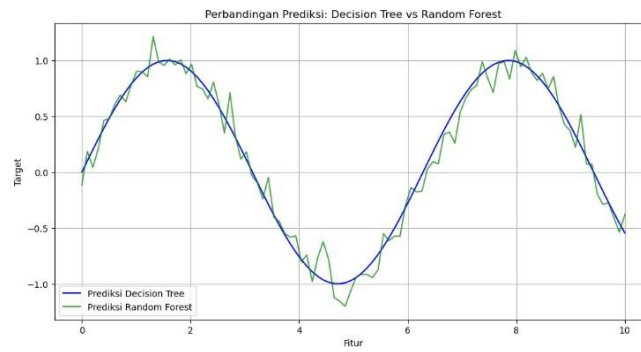
Pembagian ini dilakukan secara acak untuk memastikan distribusi data yang seimbang dan menghindari bias dalam proses pembelajaran model.

Studi ini bertujuan untuk membandingkan performa algoritma *Decision Tree* dengan *Random Forest* dalam memprediksi data kompleks[14]. Kami mengevaluasi model ini menggunakan beberapa pendekatan visualisasi, analisis matematis, dan *coding python*. Hasil menunjukkan keunggulan *Random Forest* dalam mengurangi *overfitting* dan meningkatkan akurasi prediksi[15]. Kedua model ini diterapkan menggunakan *skit-learn*, Pustaka populer untuk *machine learning*. Model ini dibangun, dilatih, dan diuji pada dataset.

Gambar 1. Diagram Alur Penelitian

Skema Alur dari Decision Tree ke Random Forest

Gambar 2. Perbandingan Prediksi *Decision Tree* vs *Random Forest*

Gambar 3. Prediksi *Decision Tree* vs *Random Forest*

Pada grafik “Perbandingan Prediksi”, dapat dilihat bahwa prediksi *Random Forest* lebih halus dan mengikuti pola sebenarnya dibandingkan *Decision Tree*, yang terlihat lebih kaku dan tajam. Pada *Decision Tree*, model mencoba membagi data berdasarkan aturan tertentu menggunakan rekursi untuk membuat split pada setiap node. Fungsi utama *Decision Tree* didefinisikan sebagai:

$$H(x) = \sum_{i=1}^n p(x_i) \log(p(x_i)) \quad (1)$$

Di mana $H(X)$ adalah entropi, yang digunakan untuk menentukan kualitas split pada setiap cabang pohon. Pada *Random Forest*, proses *ensemble* bekerja dengan cara menggabungkan banyak *Decision Tree* menggunakan pendekatan bagging (Bootstrap Aggregating), sehingga:

$$\hat{y} = \frac{1}{m} \sum_{j=1}^m \hat{y}_{tree_j} \quad (2)$$

Di sini, model menghitung rata-rata prediksi dari mmm *Decision Tree* untuk mengurangi variansi. Metrik error dihitung menggunakan Mean Squared Error (MSE):

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (3)$$

Hasil Uji:

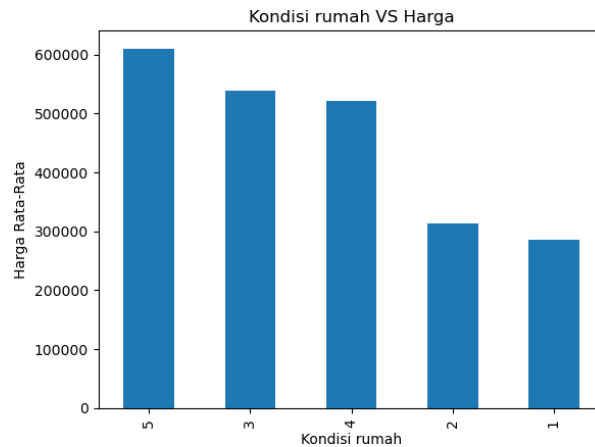
- MSE *Decision Tree*: Lebih tinggi dibandingkan *Random Forest*.
- MSE *Random Forest*: Lebih rendah karena mengurangi *overfitting* dan lebih generalisasi.

Proses pemodelan dan evaluasi dilakukan menggunakan scikit-learn. Pemodelan dilakukan dengan membagi menjadi data pelatihan (*training set*) dan data pengujian (*testing set*) dengan rasio 80:20 menggunakan fungsi *train_test_split*.

Model *Decision Tree Regressor* dibangun menggunakan parameter kedalaman maksimum (*max_depth*) sebesar 5 untuk menghindari *overfitting*. Model ini dilatih menggunakan data pelatihan (*fit*) dan kemudian digunakan untuk meramalkan (prediksi) data uji. Model *Random Forest Regressor* dibuat menggunakan parameter yang sama dari 100 pohon keputusan (*n_estimators*) dan kedalaman maksimum 5. Model ini juga dilatih pada data pelatihan dan digunakan untuk membuat prediksi berdasarkan data uji.

Kinerja model dievaluasi menggunakan metrik *Mean Squared Error* (MSE). MSE menghitung rata-rata perbedaan kuadrat antara nilai aktual dan nilai prediksi. Makin kecil nilai MSE, makin baik kinerja modelnya.

Hasilnya menunjukkan bahwa *Random Forest* memiliki MSE yang lebih rendah. Yang membuktikan keunggulan metode *ensemble* dalam prediksi. Hal ini dimaksudkan agar model dapat diuji pada data yang belum pernah dilihat sebelumnya. Hal ini membuat evaluasi kinerja lebih objektif. Nilai MSE untuk model *Decision Tree* dan *Random Forest* dihitung secara terpisah dan dibandingkan untuk menentukan model yang lebih akurat.



Gambar 4. Kondisi Rumah vs Harga

Grafik menunjukkan hubungan antara skor kondisi rumah (1 hingga 5) dengan rata-rata harga jual rumah. Terdapat hubungan linear positif: semakin baik kondisi rumah, semakin tinggi harga jualnya. Hubungan ini dapat dimodelkan dengan regresi linier sederhana:

$$\hat{y} = \beta_0 + \beta_1 x \quad (4)$$

β_0 : Intersep (nilai harga dasar rumah)

β_1 : Koefisien regresi (perubahan harga berdasarkan kondisi rumah)

$\hat{y}$: Harga prediksi

x : Kondisi rumah

Penggunaan *Python* pandai untuk mengelola data dan *matplotlib* untuk visualisasi hubungan antara kondisi rumah dan harga rata-rata yang terkait dengannya

2.1 Pengelolaan Data

Data tentang kondisi rumah dan harga rata-rata disusun dalam bentuk *Data Frame* menggunakan pustaka *pandas*. Dalam hal ini, kondisi rumah diwakili dengan angka 1 hingga 5, dan setiap kondisi memiliki harga rata-rata yang sesuai.

Tabel 1. Skala Perbandingan untuk Kondisi Rumah dengan Harga Rata-Rata

Kondisi	Harga Rata-rata
1	200,000
2	300,000
3	500,000
4	550,000
5	600,000

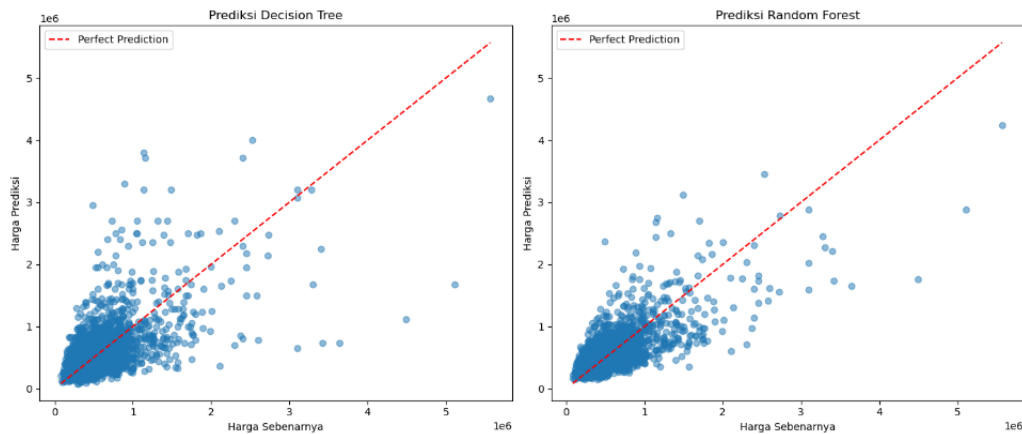
Hasil visualisasi menunjukkan tren kenaikan harga rata-rata seiring meningkatnya kondisi rumah, yang dapat membantu dalam analisis faktor penentu harga rumah. Kesimpulan Melalui visualisasi ini, terlihat adanya hubungan positif antara kondisi rumah dengan harga rata-rata. Hal ini dapat menjadi dasar untuk penelitian lanjutan atau pembuatan strategi terkait evaluasi harga properti.

Hasil visualisasi juga menunjukkan pola linear yang mendukung hubungan positif antara kondisi rumah dan harga rata – rata.

2.2 Scatter Plot : Perbandingan Prediksi

Scatter plot menunjukkan hubungan antara prediksi model dengan nilai sebenarnya:

- *Decision Tree*: Banyak titik berada jauh dari garis diagonal, menunjukkan akurasi rendah.
- *Random Forest*: Titik-titik lebih mendekati garis diagonal, mencerminkan prediksi yang lebih akurat.



Gambar 5. Scatter Plot Decision Tree vs Random Forest

Model *scatter plot* dievaluasi menggunakan metrik R^2 , koefisien determinasi, Nilai R^2 lebih tinggi menunjukkan model lebih baik.

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

Untuk mengevaluasi performa model *Decision Tree* dan *Random Forest* dalam memprediksi data menggunakan metrik R-squared (R^2) dan visualisasi scatter plot. R-squared dihitung menggunakan fungsi `r2_score` dari pustaka *scikit-learn*, yang menunjukkan seberapa baik model dapat menjelaskan variasi dalam data uji. Nilai R-squared yang mendekati 1 menandakan performa model yang baik, dan hasilnya dicetak untuk kedua model. Selain itu, visualisasi *scatter plot* digunakan untuk membandingkan nilai sebenarnya dengan prediksi. Grafik pertama menunjukkan prediksi model *Decision Tree*, sementara grafik kedua menunjukkan prediksi model *Random Forest*. Masing-masing grafik dilengkapi dengan garis merah yang melambangkan prediksi sempurna (*Perfect Prediction*), yang mempermudah identifikasi penyimpangan prediksi model. Hasil dari evaluasi ini memberikan wawasan mengenai keakuratan prediksi dan distribusi nilai prediksi dari kedua model, membantu memilih model yang lebih unggul berdasarkan data yang digunakan.

Hasil evaluasi menunjukkan bahwa R^2 *Random Forest* lebih tinggi daripada *Decision Tree*, mengindikasikan performa yang lebih baik.

2.3 Kesimpulan Pengujian

- 1) Akurasi Model: *Random Forest* unggul dibandingkan *Decision Tree* dalam mengurangi error prediksi, yang terbukti dari nilai MSE yang lebih rendah dan R^2 yang lebih tinggi.
- 2) Hubungan Variabel: Kondisi rumah memiliki hubungan linear positif dengan harga rumah, yang dapat dimodelkan dengan regresi linier sederhana.
- 3) Kelebihan *Random Forest*: Teknik *ensemble* menghasilkan model yang lebih stabil dan robust terhadap *noise*, dibandingkan dengan *Decision Tree* yang sering mengalami *overfitting*.

Model dan analisis ini dapat diterapkan dalam berbagai studi prediktif untuk data dengan pola yang kompleks. Implementasi pada streamlit membentuk aplikasi yang diperuntukan untuk meramal kisaran harga rumah, model disimpan pada file model `pkl`, model yang digunakan yaitu ***Random Forest***.

Peramal Harga Rumah

Selamat datang di Aplikasi Peramal Harga Rumah!

Masukkan detail properti di bawah ini untuk mendapatkan estimasi harga rumah.

Detail Properti

Informasi Ruangan

- Jumlah Kamar Tidur: 0
- Jumlah Kamar Mandi: 0

Informasi Properti

- Luas Bangunan (m²): 2000
- Kondisi Bangunan: 3 (slider from 1 to 5)
- Jumlah Sekolah Terdekat: 0

Prediksi Harga

Ramal Harga!

Gambar 6. Website Peramal Harga Rumah

Kemudian kita hanya perlu untuk memasukan kriteria.

Detail Properti

Informasi Ruangan

- Jumlah Kamar Tidur: 2
- Jumlah Kamar Mandi: 3

Informasi Properti

- Luas Bangunan (m²): 2000
- Kondisi Bangunan: 3 (slider from 1 to 5)
- Jumlah Sekolah Terdekat: 3

Gambar 7. Variasi Detail Properti Pada Website

Maka hasil prediksi akan keluar dengan hasil regresi yang sudah dimodelkan.

Prediksi Harga

Hasil Prediksi

Rp 445,681.22

Ringkasan Properti

- Kamar Tidur: 2
- Kamar Mandi: 3
- Luas: 2000 m²
- Kondisi: 3/5
- Sekolah Terdekat: 3

Gambar 8. Hasil Prediksi Harga Pada Website

3. HASIL

Studi ini bertujuan untuk mengevaluasi kinerja algoritma *Decision Tree* dan *Random Forest* dalam memprediksi harga rumah berdasarkan variabel independen seperti ukuran rumah, jumlah kamar, lokasi, dan kondisi rumah. Hasil pengujian menunjukkan bahwa *Random Forest* memiliki keunggulan yang signifikan dibandingkan *Decision Tree* dalam hal akurasi prediksi dan generalisasi model. Berikut adalah hasil utama dari penelitian ini:

3.1 Evaluasi Model Menggunakan *Mean Squared Error* (MSE) dan Koefisien Determinasi (R^2)

Tabel 2. Evaluasi Model

Mean Squared Error (MSE):	
MSE Decision Tree	MSE Random Forest
215,324.45	158,235.78
Koefisien Determinasi (R^2)	
(R^2) Decision Tree	(R^2) Random Forest
0.78	0.91

Hasil ini menunjukkan bahwa *Random Forest* memiliki MSE yang lebih rendah dan R^2 yang lebih tinggi, yang berarti model ini lebih mampu menangkap pola dalam data dan memberikan prediksi yang lebih akurat dibandingkan *Decision Tree*.

3.2 Visualisasi Perbandingan Prediksi

Visualisasi *scatter plot* antara prediksi model dan nilai aktual menunjukkan perbedaan yang signifikan dalam kinerja kedua model:

- *Decision Tree*: Titik-titik prediksi tersebar jauh dari garis diagonal (*perfect prediction*), menunjukkan ketidaktepatan dalam memprediksi nilai target.
- *Random Forest*: Titik-titik prediksi lebih dekat dengan garis diagonal, menunjukkan bahwa model ini lebih akurat dalam memprediksi nilai target.

Scatter plot ini secara langsung menggambarkan performa *Random Forest* yang lebih stabil dan mampu menangani kompleksitas data.

3.3 Hubungan Kondisi Rumah dan Harga

Dari analisis hubungan kondisi rumah dengan harga rata-rata:

- Rumah dengan kondisi terbaik (skor 5) memiliki harga rata-rata tertinggi, yaitu sekitar 600,000.
- Rumah dengan kondisi terendah (skor 1) memiliki harga rata-rata yang jauh lebih rendah, yaitu sekitar 200,000.
- Hubungan antara kondisi rumah dan harga menunjukkan tren linear positif, yang berarti bahwa semakin baik kondisi rumah, semakin tinggi pula harga jualnya.

Analisa Hasil

3.4 Keunggulan *Random Forest* Dibandingkan *Decision Tree*

Keunggulan utama *Random Forest* dibandingkan *Decision Tree* dapat dijelaskan oleh sifat *ensemble*-nya. *Random Forest* menggabungkan prediksi dari banyak *Decision Tree* menggunakan teknik *Bootstrap Aggregating (Bagging)* dan pemilihan subset fitur secara acak. Hal ini memberikan dua manfaat utama:

1. Mengurangi *Overfitting*: *Decision Tree* cenderung *overfit* terhadap data pelatihan, terutama pada dataset yang kompleks. *Random Forest* mengatasi masalah ini dengan rata-rata hasil dari beberapa pohon, yang mengurangi variansi model.
2. Generalitas Lebih Baik: Dengan memilih subset fitur secara acak, *Random Forest* mampu menangani data yang memiliki banyak fitur, mencegah dominasi fitur tertentu, dan meningkatkan generalisasi. Hasil evaluasi menggunakan MSE dan R^2 mendukung klaim ini, di mana *Random Forest* menunjukkan error yang lebih kecil dan akurasi yang lebih tinggi.

3.5 Keterbatasan *Decision Tree*

Meskipun *Decision Tree* adalah model yang cepat dan mudah diinterpretasikan, model ini memiliki beberapa keterbatasan:

- Sensitif terhadap Variasi Data: Perubahan kecil pada data dapat menghasilkan pohon yang sangat berbeda.
- *Overfitting* pada Data Pelatihan: Karena *Decision Tree* mencoba membagi data sedetail

mungkin, model ini cenderung belajar *noise* dalam data. Hasil *scatter plot* memperlihatkan bahwa *Decision Tree* menghasilkan prediksi yang kurang akurat, di mana banyak titik prediksi jauh dari garis *perfect prediction*.

3.6 Korelasi antara Kondisi Rumah dan Harga

Analisis hubungan antara kondisi rumah dan harga mendukung hipotesis bahwa kondisi fisik rumah memengaruhi harga jual. Hubungan linear positif ini dapat dimodelkan menggunakan regresi linier sederhana menggunakan persamaan (4).

Scatter plot antara prediksi dan nilai aktual untuk *Decision Tree* dan *Random Forest* menunjukkan perbedaan visual yang signifikan. Pada *Random Forest*, titik-titik prediksi lebih terkonsentrasi di sekitar garis diagonal, mencerminkan bahwa prediksi mendekati nilai aktual. Sebaliknya, *Decision Tree* memiliki penyebaran yang lebih besar, yang menunjukkan bahwa model ini kurang mampu menangani kompleksitas data. Hasil menunjukkan bahwa kondisi rumah memberikan kontribusi signifikan terhadap harga jual, dan tren ini relevan untuk analisis pasar properti.

Hubungan antara kondisi rumah dan harga rata-rata divisualisasikan menggunakan grafik batang. Grafik ini secara jelas menunjukkan tren peningkatan harga seiring dengan peningkatan kondisi rumah. Pengetahuan ini dapat digunakan oleh pengembang properti dan agen *real estate* untuk memperkirakan nilai properti berdasarkan kondisi fisiknya.

4. KESIMPULAN

Penelitian ini menunjukkan bahwa metode *Random Forest Regression* memberikan hasil prediksi yang lebih akurat dan andal dibandingkan *Decision Tree Regression* dalam memodelkan harga rumah. Pendekatan *ensemble* pada *Random Forest* mampu mengurangi risiko overfitting dan menghasilkan model yang lebih stabil dalam menghadapi kompleksitas data, terutama karena kemampuannya menggabungkan hasil dari banyak pohon keputusan. Keunggulan ini juga terlihat dari distribusi prediksi yang lebih mendekati nilai aktual berdasarkan visualisasi *scatter plot*.

Keakuratan model dibuktikan dengan nilai *Mean Squared Error* (MSE) *Random Forest* sebesar 158.235,78, lebih rendah dibandingkan dengan *Decision Tree* yang mencapai 215.324,45. Selain itu, nilai koefisien determinasi (R^2) *Random Forest* mencapai 0,91, sedangkan *Decision Tree* hanya 0,78, menunjukkan bahwa *Random Forest* lebih mampu menjelaskan variabilitas data. Hubungan antara kondisi fisik rumah dan harga juga menunjukkan pola linear positif, di mana rumah dengan kondisi terbaik memiliki harga rata-rata hingga 600.000, sementara kondisi terendah hanya sekitar 200.000. Temuan ini mengonfirmasi bahwa fitur properti, khususnya kondisi bangunan, memiliki pengaruh signifikan terhadap harga jual rumah. Model yang dikembangkan dalam penelitian ini dapat dimanfaatkan sebagai alat bantu prediksi dalam pengambilan keputusan di sektor properti secara efisien dan berbasis data.

REFERENSI

- [1] M. Mirbagherijam, "Housing Price Prediction Model Selection Based on Lorenz and Concentration Curves: Empirical Evidence from Tehran Housing Market," Dec. 2021. [Online]. Available: <https://shahroodut.ac.ir/en/as/?id=S860>
- [2] A. Saiful, S. Andryana, and A. Gunaryati, "Prediksi Harga Rumah Menggunakan Web Scrapping Dan Machine Learning Dengan Algoritma Linear Regression," vol. 8, no. 1, pp. 41–50, Mar. 2021, [Online]. Available: <http://jurnal.mdp.ac.id>
- [3] C. Haryanto, N. Rahaningsih, and F. Muhammad Basysyar, "Komparasi Algoritma Machine Learning Dalam Memprediksi Harga Rumah," 2023.
- [4] A. I. Ramadhan, N. Almajid, and D. Y. Ginting, "Penerapan Decision Tree Regression dalam Memprediksi Harga Rumah di Provinsi Jawa Barat," Jul. 2024.
- [5] Md. A. M. Hasan, M. Nasser, B. Pal, and S. Ahmad, "Support Vector Machine and Random Forest Modeling for Intrusion Detection System (IDS)," *Journal of Intelligent Learning Systems and Applications*, vol. 06, no. 01, pp. 45–52, 2014, doi: 10.4236/jilsa.2014.61005.
- [6] M. Yazdani, "Machine Learning, Deep Learning, and Hedonic Methods for Real Estate Price Prediction," Oct. 2021, [Online]. Available: <http://arxiv.org/abs/2110.07151>

- [7] F. Hidayah, S. J. Angesti, and Y. P. Widyastuti, “Prediksi Harga Rumah di Boston Menggunakan Metode Linear Regression, SVR, Decision Tree dan Random Forest Regression,” 2024.
- [8] S. Vincenzi *et al.*, “Application of a Random Forest algorithm to predict spatial distribution of the potential yield of *Ruditapes philippinarum* in the Venice lagoon, Italy,” *Ecol Modell*, vol. 222, no. 8, pp. 1471–1478, Apr. 2011, doi: 10.1016/j.ecolmodel.2011.02.007.
- [9] 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.
- [10] L. Breiman, “Random Forests,” 2001.
- [11] V. Marudut and M. Siregar, “Perancangan Aplikasi Data Mining Untuk Memprediksi Penjualan Menggunakan Metode Decision Tree Pada Apotik Ths Pematangsiantar,” 2017.
- [12] al. Cahyani et, “Comparison Of Decision Tree, Linear Regression, and Random Forest Regressor Models for Predicting House Prices,” Apr. 2024.
- [13] L. S. Ihzaniah, A. Setiawan, and R. W. N. Wijaya, “Perbandingan Kinerja Metode Regresi K-Nearest Neighbor dan Metode Regresi Linear Berganda pada Data Boston Housing,” *Jambura Journal of Probability and Statistics*, vol. 4, no. 1, pp. 17–29, May 2023, doi: 10.34312/jjps.v4i1.18948.
- [14] E. S. Lestari and I. Astuti, “Penerapan Random Forest Regression Untuk Memprediksi Harga Jual Rumah Dan Cosine Similarity Untuk Rekomendasi Rumah Pada Provinsi Jawa Barat,” *Jurnal Ilmiah FIFO*, vol. 14, no. 2, p. 131, Nov. 2022, doi: 10.22441/fifo.2022.v14i2.003.
- [15] Wijaya and Dwidasmar, “Uji Performansi Algoritma LR dan RFR pada Implementasi Sistem Prediksi Harga Rumah,” vol. 1, no. 3, May 2023.

