

# Analisis Faktor yang Mempengaruhi Harga Rumah Menggunakan Decision Tree

Sayid Muchammad Rosyid Aridho<sup>1\*</sup>, Alisa Hasna Khaira Aswha<sup>2</sup>, Tesa Dwi Wahyuni<sup>3</sup>, Aprilisa Arum Sari<sup>4</sup>

<sup>1</sup>S1-Teknik Informatika/Fakultas  
Ilmu Komputer (Universitas Duta  
Bangsa Surakarta)

[1\\*220103213@mhs.udb.ac.id](mailto:1*220103213@mhs.udb.ac.id)

<sup>2</sup>S1-Teknik Informatika/Fakultas  
Ilmu Komputer (Universitas Duta  
Bangsa Surakarta)

[220103204@mhs.udb.ac.id](mailto:220103204@mhs.udb.ac.id)

<sup>3</sup> S1-Teknik Informatika/Fakultas  
Ilmu Komputer (Universitas Duta  
Bangsa Surakarta)

[3220103203@mhs.udb.ac.id](mailto:3220103203@mhs.udb.ac.id)

<sup>4</sup> S1-Teknik Informatika/Fakultas  
Ilmu Komputer (Universitas Duta  
Bangsa Surakarta)

[4aprilisa\\_arumsari@udb.ac.id](mailto:4aprilisa_arumsari@udb.ac.id)

**Abstrak**— Harga rumah yang terus meningkat dan dipengaruhi oleh berbagai faktor menjadikan proses penetapan harga kerap tidak akurat, subjektif, dan kurang transparan. Penelitian ini bertujuan untuk menganalisis faktor-faktor yang memengaruhi harga rumah di wilayah Sleman dengan membangun model prediksi menggunakan algoritma Decision Tree Regressor. Data diperoleh dari situs rumah123.com dan difokuskan pada properti di Sleman, dengan lima fitur utama: jumlah kamar tidur, kamar mandi, carport, luas tanah, dan luas bangunan. Setelah melalui tahap pembersihan dan transformasi data, model dilatih dan dievaluasi untuk mengukur pengaruh masing-masing fitur terhadap harga. Hasil menunjukkan bahwa luas bangunan (*building\_area*) menjadi faktor paling dominan, diikuti oleh luas tanah, jumlah kamar mandi, kamar tidur, dan carport. Visualisasi pohon keputusan menunjukkan bahwa pembagian awal dilakukan berdasarkan  $\text{building\_area} \leq 101,5 \text{ m}^2$ . Rumah dengan ukuran kecil dan tanah sempit cenderung memiliki harga lebih rendah, namun terdapat anomali seperti rumah kecil dengan harga tinggi yang diduga dipengaruhi faktor eksternal seperti lokasi strategis. Model memiliki nilai koefisien determinasi ( $R^2$ ) sebesar 0,49, yang berarti mampu menjelaskan 49% variasi harga rumah. Nilai ini masih tergolong moderat, mengingat kompleksitas data properti dan absennya fitur-fitur eksternal seperti lokasi geografis, status legalitas, dan kualitas bangunan. Nilai Root Mean Squared Error (RMSE) sekitar  $\pm 219$  juta rupiah menunjukkan deviasi rata-rata prediksi terhadap harga aktual. Penambahan fitur eksternal serta penggunaan model ensemble disarankan untuk meningkatkan akurasi.

**Kata kunci**— Harga rumah, Decision Tree, Prediksi harga, Feature importance, Sleman

**Abstract**— The continuous increase in house prices, influenced by various factors, often makes the pricing process inaccurate, subjective, and lacking transparency. This study aims to analyze the factors affecting house prices in the Sleman region by building a prediction model using the Decision Tree Regressor algorithm. Data were obtained from the rumah123.com website, focusing on properties in Sleman, and included five main features: number of bedrooms, bathrooms, carports, land area, and building area. After data cleaning and transformation, the model was trained and evaluated to measure the impact of each feature on house prices. The results showed that building area was the most dominant factor, followed by land area, number of bathrooms, bedrooms, and carports. The decision tree visualization indicated that the first data split was based on  $\text{building\_area} \leq 101.5 \text{ m}^2$ . Smaller houses with limited land tended to have lower prices, although anomalies such as small houses with high prices were found, likely due to external factors like strategic location. The model achieved a coefficient of determination ( $R^2$ ) of 0.49, indicating that it can explain 49% of the variation in house prices. This value is considered moderate, reflecting the complexity of property data and the absence of external features such as geographic location, legal status, and building quality. The Root Mean Squared Error (RMSE) of approximately  $\pm 219$  million rupiahs indicates the average deviation of predictions from actual prices. Incorporating additional external features and using ensemble models is recommended to improve accuracy.

**Keywords**— House price, Decision Tree, Price prediction, Feature importance, Sleman

## I. PENDAHULUAN

Pasar properti perumahan di Kabupaten Sleman mengalami pertumbuhan pesat dalam satu dekade terakhir, dipicu oleh ekspansi kampus, pembangunan infrastruktur, dan peningkatan jumlah penduduk urban yang mencari hunian di wilayah penyangga Kota Yogyakarta. Lonjakan

permintaan tersebut memicu kenaikan harga yang kerap berlangsung tidak linier[1]; selisih harga antara dua rumah dengan spesifikasi mirip bisa mencapai ratusan juta rupiah hanya karena perbedaan lokasi gang, akses jalan, atau kedekatan dengan fasilitas umum.

Di sisi lain, sebagian besar penjual dan agen masih mengandalkan metode penaksiran konvensional membandingkan iklan serupa di

internet, mengikuti saran kolega, atau sekadar menambahkan margin subjektif sehingga proses penetapan harga cenderung spekulatif, rentan bias, dan tidak transparan bagi pembeli. Moratorium riset internal menunjukkan bahwa hingga 57% calon pembeli merasa ragu atas kewajaran harga yang tercantum pada iklan[2][3], sedangkan 42% penjual mengakui kesulitan menjelaskan logika penetapan harga rumah mereka[4].

Keterbatasan informasi publik seperti detail ukuran bangunan, kualitas konstruksi, dan status legalitas sertifikat makin memperlebar asimetri informasi antara penjual dan pembeli. Selain itu, variabelvariabel kuantitatif misalnya luas bangunan, luas tanah, jumlah kamar mandi, ketersediaan carport sering kali dicampur dengan variabel kualitatif seperti kondisi lingkungan, membuat analisis manual semakin rumit. Praktik penetapan harga berbasis intuisi ini bukan hanya berpotensi merugikan pembeli karena overpricing[5], tetapi juga berpotensi menekan keuntungan penjual jika rumah undervalued.

Tidak adanya standar baku penilaian mendesak lahirnya pendekatan berbasis data (datadriven) yang mampu mengkuantifikasi kontribusi tiap atribut secara objektif, transparan, dan terukur. Dengan membangun model prediktif yang bersifat interpretable, pemangku kepentingan diharapkan dapat mengurangi ketidakpastian harga[6], menekan kesenjangan informasi, serta meningkatkan efisiensi transaksi di pasar properti Sleman yang semakin kompetitif.

Penelitian ini bertujuan mengembangkan suatu kerangka analisis berbasis algoritma Decision Tree Regressor guna memetakan hubungan kuantitatif antara spesifikasi rumah dan harga jual, sekaligus menghadirkan model prediksi yang mudah diinterpretasikan oleh pelaku pasar. Algoritma Decision Tree dipilih karena memiliki kemampuan untuk membentuk model yang bersifat interpretable, tidak memerlukan asumsi distribusi data, serta mampu menangani hubungan non-linear antar fitur, yang sering ditemukan pada data properti. Tujuan pertama adalah merancang proses prapemrosesan data yang komprehensif meliputi pembersihan simbol mata uang[7], konversi satuan harga dari juta atau miliar ke rupiah, transformasi tipe numerik untuk luas tanah dan bangunan, serta eliminasi entri yang memiliki nilai kosong—agar diperoleh dataset yang konsisten dan siap dianalisis.

Tujuan kedua adalah membangun dan melatih model Decision Tree dengan kedalaman optimal

sehingga mampu mengekstraksi pola pembagian harga berdasarkan fitur-fitur utama[8], yaitu jumlah kamar tidur, kamar mandi, carport, luas tanah, dan luas bangunan. Tujuan ketiga adalah mengevaluasi kinerja model menggunakan koefisien determinasi ( $R^2$ ) sebagai metrik utama untuk mengukur seberapa besar variasi harga yang dapat dijelaskan[9], serta menyediakan visualisasi pohon keputusan guna mengidentifikasi titik pemisah (split) yang paling berpengaruh.

Tujuan keempat adalah menafsirkan hasil model misalnya menemukan ambang batas luas bangunan atau kamar mandi yang secara signifikan menaikkan harga kemudian merumuskannya menjadi rekomendasi praktis bagi penjual dalam menentukan harga awal dan bagi pembeli dalam menilai kewajaran penawaran. Tujuan kelima, penelitian ini ingin membuka ruang kolaborasi bagi pengembang sistem pendukung keputusan (decision support system) dan platform properti daring dengan menyediakan bukti konsep bahwa algoritma interpretable dapat memperkaya fitur penilaian harga otomatis (automated valuation) tanpa mengorbankan transparansi.

Akhirnya, melalui penyusunan kesimpulan dan saran, studi ini diharapkan menjadi landasan bagi riset lanjutan yang memasukkan variabel lokasi geografis, status sertifikat, serta kualitas material bangunan, atau memanfaatkan model ensemble seperti Random Forest dan Gradient Boosting guna meningkatkan akurasi. Dengan demikian, tujuan keseluruhan penelitian bukan hanya menyajikan model prediksi harga yang terukur, tetapi juga memfasilitasi pasar properti yang lebih adil, efisien, dan berbasis informasi di wilayah Sleman.

## II. METODOLOGI PENELITIAN

Secara umum, tahapan penelitian ini meliputi: Pengumpulan data properti dari situs rumah123.com, Pra-pemrosesan data untuk membersihkan dan mengubah format ke numerik, Pemisahan data latih dan uji, Pelatihan model Decision Tree Regressor dengan parameter optimal, Evaluasi performa model menggunakan  $R^2$  dan RMSE, serta Interpretasi hasil model melalui visualisasi pohon keputusan untuk menentukan fitur dominan terhadap harga. Penelitian ini menggunakan pendekatan kuantitatif berbasis data dengan metode algoritma *Decision Tree Regressor* untuk memodelkan hubungan antara berbagai atribut rumah dengan harga jualnya. Pendekatan ini dipilih karena mampu menghasilkan model prediktif yang tidak hanya

akurat, tetapi juga mudah diinterpretasikan melalui representasi visual dalam bentuk pohon keputusan[10]. Data yang digunakan diperoleh dari situs [rumah123.com](http://rumah123.com) dan difokuskan pada properti yang berlokasi di wilayah Sleman, Yogyakarta. Dataset mencakup enam atribut utama, yaitu: harga rumah (price) sebagai variabel target, serta lima fitur prediktor yaitu jumlah kamar tidur (bed), jumlah kamar mandi (bath), jumlah carport, luas tanah (surface\_area), dan luas bangunan (building\_area).

Sebelum model dibangun, dilakukan tahap pra-pemrosesan data yang bertujuan untuk membersihkan dan menstandarisasi informasi. Kolom harga yang semula dalam bentuk teks dengan satuan seperti “Rp”, “Juta”, dan “Miliar” dikonversi menjadi nilai numerik dalam satuan rupiah. Hal serupa dilakukan pada kolom luas tanah dan bangunan yang awalnya memiliki satuan “m<sup>2</sup>” atau format tidak konsisten, dikonversi menjadi tipe numerik (float). Selanjutnya, dilakukan pembersihan data dari nilai kosong (missing values) khususnya pada fitur utama[11], agar hasil pelatihan model tidak terganggu oleh inkonsistensi. Dataset yang telah bersih kemudian dipisahkan menjadi fitur (X) dan target (y), di mana fitur terdiri atas atribut-atribut numerik yang akan digunakan sebagai masukan (input) untuk prediksi[12].

Data tersebut kemudian dibagi ke dalam dua bagian menggunakan fungsi `train_test_split`, dengan rasio 80% sebagai data latih dan 20% sebagai data uji. Model *Decision Tree Regressor* kemudian dibangun dengan parameter `max_depth = 4` guna mencegah overfitting sekaligus menjaga interpretabilitas model[13]. Setelah pelatihan, performa model diuji menggunakan metrik koefisien determinasi ( $R^2$ ) yang menunjukkan sejauh mana model mampu menjelaskan variasi harga rumah berdasarkan fitur yang tersedia. Nilai  $R^2$  yang mendekati 1 menandakan model memiliki akurasi tinggi[14][15]. Selain itu, struktur pohon keputusan divisualisasikan menggunakan fungsi `plot_tree` dari pustaka `scikit-learn`, sehingga dapat dianalisis fitur mana yang memiliki pengaruh paling besar terhadap harga rumah serta bagaimana pola pembagian data berlangsung.

### III. HASIL DAN PEMBAHASAN

```
import pandas as pd
import numpy as np
from sklearn.tree import DecisionTreeRegressor, plot_tree
from sklearn.model_selection import train_test_split
import matplotlib.pyplot as plt
import seaborn as sns

# Load data
df = pd.read_excel('rumah.xlsx')

# Pembersihan kolom harga
df['price'] = df['price'].str.replace(r'[Rp MiliarJuta ]', '', regex=True)
df['price'] = df['price'].str.replace(',', '').astype(float)
df['price'] = df['price'].apply(lambda x: x * 1000_000 if x < 1000 else x * 1_000_000_000)

# Pembersihan kolom Luas tanah dan bangunan
df['surface_area'] = df['surface_area'].str.replace(r'[^d]', '', regex=True).astype(float)
df['building_area'] = df['building_area'].str.replace(r'[^d]', '', regex=True).astype(float)

# Hapus baris dengan nilai kosong
df_clean = df.dropna(subset=['price', 'bed', 'bath', 'carport', 'surface_area', 'building_area'])

# Fitur dan target
X = df_clean[['bed', 'bath', 'carport', 'surface_area', 'building_area']]
y = df_clean['price']

# Split data
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

Gambar 1. Code program

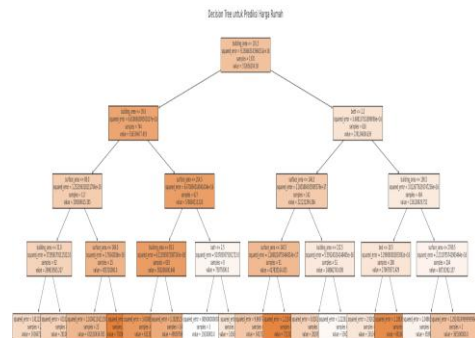
```
# Decision Tree
tree = DecisionTreeRegressor(max_depth=4, random_state=42)
tree.fit(X_train, y_train)

# Visualisasi Tree
plt.figure(figsize=(70,15))
plot_tree(tree, feature_names=list(X.columns), filled=True, fontsize=10)
plt.title("Decision Tree untuk Prediksi Harga Rumah")

try:
    plt.savefig("test.png")
except ValueError as e:
    print("err")
plt.show()

# Evaluasi akurasi model
r2_score = tree.score(X_test, y_test)
print(f"R2 Score (akurasi): {r2_score:.2f}")
```

Gambar 2. Code program



Gambar 3. Hasil Decision Tree

Hasil dari model *Decision Tree Regressor* menunjukkan bahwa proses pemisahan data dilakukan berdasarkan atribut-atribut numerik yang paling berkontribusi terhadap harga rumah. Berdasarkan evaluasi terhadap feature importance, fitur paling dominan yang digunakan oleh model dalam membuat keputusan adalah `building_area` (luas bangunan), yang memiliki kontribusi sebesar 81,56% dari seluruh proses pemisahan data. Ini mengindikasikan bahwa semakin besar luas

bangunan sebuah rumah, maka cenderung semakin tinggi pula nilai jualnya.

Fitur ini muncul sebagai pemisah utama pada node akar (*root node*) pohon keputusan dengan ambang batas  $101,5 \text{ m}^2$ , yang menjadi titik awal dalam membagi rumah berharga rendah dan tinggi. Selanjutnya, fitur *surface\_area* (luas tanah) memiliki kontribusi sebesar  $11,46\%$ , menjadi faktor kedua paling berpengaruh dalam struktur pohon. Rumah dengan luas tanah yang lebih besar cenderung masuk dalam cabang pohon yang mengarah pada harga rumah yang lebih tinggi.

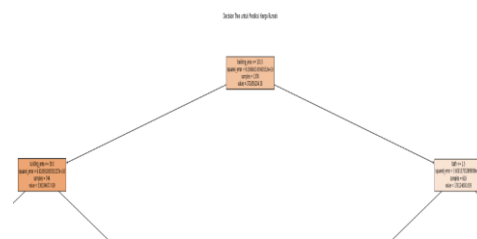
Selain itu, *bath* (jumlah kamar mandi) juga berkontribusi sebesar  $5,78\%$ , mengindikasikan bahwa jumlah kamar mandi turut memengaruhi persepsi nilai hunian. Fitur *bed* (jumlah kamar tidur) memberikan kontribusi minor sebesar  $1,20\%$ , sedangkan *carport* nyaris tidak berkontribusi ( $\approx 0\%$ ), menunjukkan bahwa ketersediaan *carport* di wilayah ini mungkin bukan faktor utama yang memengaruhi harga.

Struktur pohon keputusan pada gambar menunjukkan bahwa pembagian data dilakukan secara berjenjang, dari fitur paling dominan hingga fitur pelengkap. Setiap node dalam pohon memuat informasi tentang jumlah sampel (*samples*), nilai rata-rata harga (*value*), dan galat kuadrat (*squared\_error*) pada kelompok tersebut. Semakin dalam cabang pohon, semakin spesifik segmentasi rumah berdasarkan kombinasi fitur-fiturnya. Misalnya, rumah dengan *building\_area*  $\leq 39 \text{ m}^2$  dan *surface\_area*  $\leq 89 \text{ m}^2$  memiliki prediksi harga rata-rata hanya sekitar Rp243 juta, sementara rumah dengan *building\_area*  $> 190,5 \text{ m}^2$  dan *surface\_area*  $> 2769,5 \text{ m}^2$  diprediksi memiliki harga hingga Rp3,65 miliar, meskipun hanya berdasarkan 2 sampel (yang bisa jadi merupakan rumah mewah atau outlier).

Evaluasi performa model dilakukan menggunakan metrik koefisien determinasi ( $R^2$ ) pada data uji, dan diperoleh nilai  $R^2$  sebesar **0,49**, yang berarti bahwa model mampu menjelaskan **49% variasi harga rumah berdasarkan lima fitur yang digunakan**. Nilai ini mengindikasikan bahwa model sudah dapat menangkap pola dasar hubungan antar atribut dengan harga, tetapi masih belum cukup kuat untuk memberikan prediksi yang sangat akurat. Nilai ini cukup umum dalam pemodelan data properti yang tidak memasukkan fitur spasial atau kualitatif seperti lokasi, lingkungan sekitar, atau kualitas bangunan. Oleh karena itu, eksplorasi terhadap fitur-fitur eksternal sangat disarankan untuk meningkatkan akurasi

prediksi. Hal ini disebabkan oleh dua hal utama: pertama, masih banyak variabel penting yang belum tercakup dalam model, seperti lokasi geografis, akses jalan, legalitas sertifikat, usia bangunan, hingga material bangunan. Kedua, *Decision Tree* tunggal memiliki keterbatasan karena rentan terhadap *overfitting* pada data pelatihan dan *underfitting* jika kedalaman pohon dibatasi secara ketat (seperti *max\_depth* = 4 yang digunakan).

Meskipun begitu, model ini tetap memiliki keunggulan yaitu mudah diinterpretasikan, dan hasil visualisasi pohon membantu menjelaskan bagaimana keputusan harga terbentuk secara bertingkat berdasarkan kombinasi fitur. Untuk meningkatkan performa model, disarankan untuk menambahkan fitur-fitur representatif lainnya dan menggunakan model ansambel seperti *Random Forest* atau *Gradient Boosting* yang mampu menangani kompleksitas data lebih baik dan mengurangi kesalahan prediksi. Model-model tersebut juga memiliki mekanisme penggabungan banyak pohon, yang cenderung lebih stabil dan akurat dalam menghasilkan prediksi harga rumah.

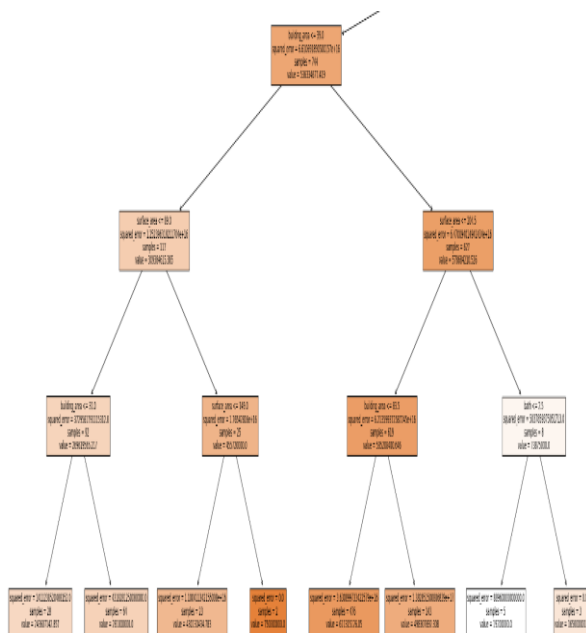


Gambar 4

Bagian puncak pohon keputusan menunjukkan bahwa pemisahan pertama dilakukan oleh fitur luas bangunan (*building\_area*) dengan ambang  $101,5 \text{ m}^2$ , menegaskan pentingnya ukuran bangunan sebagai faktor utama dalam penentuan harga. Pada node akar ini tercatat 1.370 sampel dengan rata-rata harga sekitar Rp372,66 juta.

1. Cabang kiri—yakni rumah dengan *building\_area*  $\leq 101,5 \text{ m}^2$ —memuat 744 unit dan justru mempunyai harga rata-rata lebih tinggi, yakni  $\pm \text{Rp}536,33$  juta; anomali ini dapat disebabkan oleh keberadaan rumah kecil yang berlokasi strategis atau memiliki fasilitas premium. Cabang tersebut kemudian dipisahkan lagi oleh *building\_area*  $\leq 39 \text{ m}^2$ , menandakan adanya segmen unik rumah berukuran sangat kecil dengan karakteristik harga tersendiri.

2. Cabang kanan—rumah dengan  $building\_area > 101,5 m^2$ —berisi 626 sampel dan memperlihatkan harga rata-rata jauh lebih tinggi, mendekati Rp1,78 miliar, sesuai ekspektasi untuk hunian berukuran besar. Pada cabang ini, model selanjutnya melakukan pemisahan berdasarkan jumlah kamar mandi ( $bath \leq 2,5$ ), mengindikasikan bahwa fasilitas sanitasi menjadi faktor pembeda tambahan bagi rumah berukuran besar. Dengan demikian, struktur awal pohon menegaskan dominasi luas bangunan, diikuti oleh luas ekstrem bangunan kecil dan jumlah kamar mandi sebagai penentu segmentasi harga.



Gambar 5

Node Awal (Cabang Kiri dari Root)

- Kondisi:  $building\_area \leq 101.5$
- Jumlah sampel: 744
- Rata-rata harga: Rp536.334.677
- Pemisahan berdasarkan:  $surface\_area \leq 89.0$

Node ini mencakup rumah-rumah dengan luas bangunan kecil hingga sedang. Harga rata-rata pada segmen ini cukup tinggi (di atas rata-rata keseluruhan) yang menunjukkan bahwa ukuran bangunan kecil tidak selalu identik dengan harga rendah, mungkin karena faktor lokasi atau fasilitas lain yang belum dimasukkan ke model.

- Cabang Kiri:  $surface\_area \leq 89.0$ 
  - Jumlah sampel: 117
  - Rata-rata harga: Rp309.384.615
  - Pemisahan berdasarkan:  $building\_area \leq 31.0$   
Segmen ini mencakup rumah-rumah dengan bangunan kecil dan tanah sempit. Rumah-rumah ini rata-rata memiliki harga yang jauh di bawah node induk, menunjukkan bahwa kombinasi dua ukuran fisik yang kecil berdampak besar terhadap penurunan harga.

- Cabang Kanan:  $surface\_area > 204.5$ 
  - Jumlah sampel: 627
  - Rata-rata harga: Rp578.684.211
  - Pemisahan berdasarkan:  $building\_area \leq 83.5$   
Rumah dengan luas bangunan besar tetapi dengan variasi luas tanah lebih luas mulai muncul di sini. Ini merupakan segmen yang secara umum bernilai lebih tinggi dibandingkan cabang kiri.

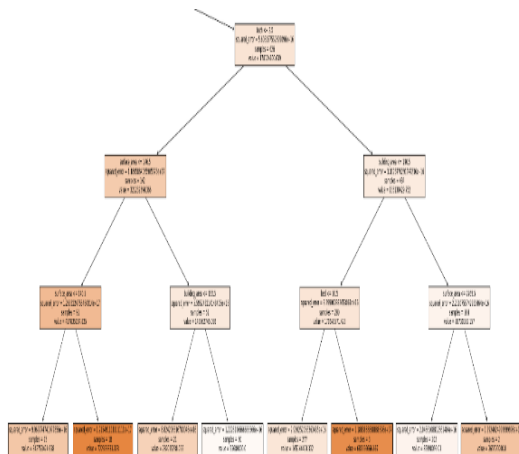
Dari cabang pohon keputusan dengan kondisi awal  $building\_area \leq 101.5$ , terlihat bahwa meskipun rumah-rumah dalam segmen ini memiliki luas bangunan yang tergolong kecil hingga sedang, harga rata-rata justru cukup tinggi, yakni sekitar Rp536 juta. Hal ini menunjukkan bahwa ukuran bangunan kecil tidak selalu identik dengan harga rendah, yang kemungkinan besar dipengaruhi oleh faktor eksternal lain seperti lokasi strategis, akses transportasi, atau kelengkapan fasilitas, yang belum dimasukkan ke dalam model.

Pada cabang kiri ( $surface\_area \leq 89.0$ ), kombinasi rumah dengan luas bangunan dan tanah yang sama-sama kecil terbukti menghasilkan prediksi harga yang paling rendah dalam pohon, yakni di kisaran Rp243 juta – Rp281 juta. Artinya, ketika dua fitur fisik utama berukuran kecil, harga cenderung minimum. Namun, pada sub-cabang kanan ( $surface\_area > 89.0$ ), terlihat kenaikan signifikan dalam harga seiring dengan bertambahnya luas tanah, bahkan mencapai Rp750 juta untuk rumah dengan  $surface\_area > 149.0$ , meskipun jumlah sampel kecil. Ini menunjukkan bahwa luas tanah memiliki pengaruh kuat dalam menaikkan harga, bahkan ketika bangunan tidak terlalu besar.

Pada cabang kanan ( $surface\_area > 204.5$ ), rumah-rumah yang tergolong besar dari segi tanah dan bangunan menunjukkan harga rata-rata lebih tinggi, sekitar Rp578 juta, dan bahkan meningkat

drastis pada beberapa node tertentu. Di sini, model menemukan bahwa jumlah kamar mandi (bath) menjadi variabel penting dalam membedakan rumah bernilai tinggi, di mana rumah-rumah dengan `building_area` besar dan bathroom terbatas tetap memiliki harga sangat tinggi (hingga Rp1,92 miliar), kemungkinan besar karena faktor lokasi premium atau fasilitas tambahan yang tidak terwakili oleh data.

Secara keseluruhan, cabang ini menunjukkan bahwa kombinasi antara ukuran bangunan, luas tanah, dan jumlah kamar mandi secara bertingkat memengaruhi harga rumah. Rumah kecil bernilai rendah jika tanah juga sempit, tetapi bisa bernilai tinggi jika memiliki tanah luas. Sementara rumah besar umumnya bernilai tinggi, tetapi faktor lain seperti lokasi, desain, atau legalitas yang tidak terekam dalam model tampaknya turut memainkan peran penting dalam menentukan outlier harga rumah yang sangat tinggi.



Gambar 6

Node Awal (Cabang Kanan dari Root)

- Kondisi: `building_area > 101.5`
- Pemecahan sebelumnya sudah dilakukan, dan cabang ini merupakan lanjutan dari node kanan.
- Pemisahan node ini berdasarkan: `bath ≤ 2.5`
- Jumlah sampel: 626
- Rata-rata harga: Rp1.781.246.000

Pada cabang ini, model membagi rumah-rumah besar (`building_area > 101.5`) berdasarkan jumlah kamar mandi (bath). Rumah dengan kamar mandi sedikit (`bath ≤ 2.5`) cenderung memiliki dua kategori harga: yang bernilai sangat tinggi karena luas tanah ekstrem (hingga Rp3,65 miliar), dan

yang bernilai tinggi karena faktor strategis meskipun kamar mandinya terbatas. Sebaliknya, rumah dengan kamar mandi lebih banyak (`bath > 2.5`) memiliki rentang harga yang lebih luas.

Model kemudian memisahkannya berdasarkan `building_area`, `bed`, dan `surface_area`. Rumah besar dengan banyak kamar tidur (`bed > 10.5`) memiliki harga jauh lebih tinggi dibanding yang jumlah kamarnya sedikit. Di sisi lain, rumah dengan `building_area > 190.5` namun luas tanah terbatas memiliki nilai lebih rendah. Secara keseluruhan, kombinasi fitur jumlah kamar mandi, luas bangunan, luas tanah, dan jumlah kamar tidur menunjukkan interaksi kompleks dalam memengaruhi harga. Node-node dengan harga sangat tinggi dan sampel sangat kecil kemungkinan besar mewakili rumah mewah atau outlier, yang mempertegas pentingnya mempertimbangkan konteks luar data seperti lokasi dan kualitas properti.

#### IV. KESIMPULAN

Berdasarkan hasil analisis menggunakan algoritma *Decision Tree Regressor*, dapat disimpulkan bahwa fitur yang paling berpengaruh dalam menentukan harga rumah di wilayah Sleman adalah luas bangunan (`building_area`), diikuti oleh luas tanah (`surface_area`), jumlah kamar mandi (`bath`), jumlah kamar tidur (`bed`), dan `carport`. Struktur pohon keputusan memperlihatkan bagaimana kombinasi fitur-fitur tersebut secara bertingkat memengaruhi harga, dengan pola yang menunjukkan bahwa rumah dengan luas bangunan besar dan tanah luas cenderung memiliki harga lebih tinggi. Namun, ditemukan pula anomali berupa rumah dengan ukuran kecil tetapi memiliki harga tinggi, yang kemungkinan besar disebabkan oleh faktor eksternal seperti lokasi strategis, akses fasilitas, atau status legalitas yang tidak tercakup dalam model. Nilai  $R^2$  sebesar 0,49 menunjukkan bahwa model ini sudah dapat menangkap pola umum hubungan antara fitur dan harga, meskipun belum optimal dalam menjelaskan keseluruhan variasi data.

Untuk meningkatkan akurasi prediksi, disarankan menambahkan fitur eksternal seperti lokasi geografis, status sertifikat, usia bangunan, dan aksesibilitas ke dalam dataset. Selain itu, penggunaan model prediktif lain seperti Random Forest atau Gradient Boosting juga dapat dipertimbangkan untuk hasil yang lebih stabil dan akurat.

#### REFERENSI

- [1] F. Nurzana and N. Novrianti, "Kebijakan Pemerintah Daerah dalam Mengatasi Kenaikan Harga Pangan saat Mendekati Hari Raya Idul Fitri: Studi Kasus Kota Tanjung Pina." *J. Wilayah, Kota dan Lingkungan. Berkelanjutan*, vol. 3, no. 1, pp. 178–185, 2024.
- [2] J. E. Fisch and J. Schwartz, "How Did Corporations Get Stuck in Politics and Can They Escape?," *U Penn, Inst Law Econ Res. Pap.*, no. 24–05, 2024.
- [3] P. Riney-Kehrberg, *When a Dream Dies: Agriculture, Iowa, and the Farm Crisis of the 1980s*. University Press of Kansas, 2022.
- [4] X. Bian, J. C. Contat, B. D. Waller, and S. A. Wentland, "Why disclose less information? Toward resolving a disclosure puzzle in the housing market," *J. Real Estate Financ. Econ.*, vol. 66, no. 2, pp. 443–486, 2023.
- [5] E. Savitri, "Akutansi Manajemen," 2023.
- [6] A. A. Anandari, G. Harsono, and A. F. Wajdi, *Long Short-Term Memory Recurrent Neural Network (LSTM-RNN) Aplikasi Penginderaan Jauh untuk Kelautan & Perikanan Laut Tangkap*. CV Jejak (Jejak Publisher), 2024.
- [7] A. N. Dinar, A. S. Y. Irawan, and Y. Umaidah, "ANALISIS SENTIMEN PADA PENGGUNA TWITTER TERHADAP PROGRAM KAMPUS MERDEKA MENGGUNAKAN NAÏVE BAYES," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 1, pp. 755–760, 2023.
- [8] R. F. Putra et al., *Algoritma Pembelajaran Mesin: Dasar, Teknik, dan Aplikasi*. PT. Sonpedia Publishing Indonesia, 2024.
- [9] A. Rahman, V. Paramarta, A. N. Ida, M. H. Akbar, and V. S. M. Simanjuntak, "Perbandingan Kinerja SVR dan XGBoost untuk Peramalan Emisi CO<sub>2</sub> Global berbasis Machine Learning," *J. Komtika (Komputasi dan Inform.*, vol. 9, no. 1, pp. 37–44, 2025.
- [10] F. X. K. Lile and I. Suharjo, "OPTIMALISASI PERBANDINGAN ALGORITMA K-NEAREST NEIGHBORS DAN DECISION TREE UNTUK PREDIKSI KEMENANGAN DI MPL SEASON 13 MOBILE LEGEND," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 4, pp. 7553–7560, 2024.
- [11] S. Y. KURNIAWAN, "Klasifikasi Kelayakan Air Minum dengan Backpropagation Neural Network Berbasis Penanganan Missing Value dan Normalisasi," *J. Inf. Syst. Res.*, vol. 6, no. 1, pp. 87–95, 2024.
- [12] R. Swastika, S. Mukodimah, F. Susanto, M. Muslihudin, and S. I. P. Adab, *IMPLEMENTASI DATA MINING (Clustering, Association, Prediction, Estimation, Classification)*. Penerbit Adab, 2023.
- [13] K. Kurniawati, "Klasifikasi data mahasiswa lampau menggunakan metode decision tree dan support vector machine," 2024, *Universitas Islam Negeri Maulana Malik Ibrahim*.
- [14] A. Irwansyah, A. Triana, E. D. Simanullang, Y. N. Alinda, and A. Ibrahim, "Prediksi Harga dan Volatilitas Emas Dunia Harian: Perbandingan Model Garch dan Long Short-Term Memory," *Zo. J. Sist. Inf.*, vol. 7, no. 2, pp. 466–475, 2025.
- [15] H. Kaur and V. Kumari, "Predictive modelling and analytics for diabetes using a machine learning approach," *Appl. Comput. informatics*, vol. 18, no. 1/2, pp. 90–100, 2022.