



Penerapan Algoritma Linear Regression dalam Memprediksi Harga Rumah Menggunakan RapidMiner

Siti Khoiriyah^{1*}, Zaehol Fatah²¹ Teknologi Informasi, Fakultas Sains dan Teknologi, Universitas Ibrahimy, Situbondo, Jawa Timur, Indonesia² Sistem Informasi, Fakultas Sains dan Teknologi, Universitas Ibrahimy, Situbondo, Jawa Timur, Indonesia**Info Artikel****Riwayat Artikel:**Diterima : **20-10-2024**Direvisi : **28-10-2024**Publikasi : **15-12-2024****Kata Kunci:**

Algoritma Linear Regression,
Data Mining,
Prediksi harga rumah,
Variabel Dependen,
Variabel Independen

Keywords:

*Linear Regression Algorithm,
Data Mining,
House Price Prediction,
Dependent Variable,
Independent Variable*

ABSTRAK

Permintaan pasar properti di Indonesia yang terus meningkat mempengaruhi harga rumah, yang dipengaruhi oleh berbagai faktor. Klasifikasi, sebagai salah satu metode dalam data mining, memiliki kemampuan untuk melakukan prediksi. Model prediksi harga rumah dikembangkan menggunakan teknik regresi linier. Hasil analisis menunjukkan bahwa faktor utama yang memengaruhi harga rumah adalah luas tanah, jumlah kamar mandi, jumlah lantai, dan fasilitas parkir. Model ini akurat, terutama untuk harga rumah rendah hingga menengah. Faktor luas area (koefisien 242.186, p-value sangat kecil), jumlah kamar mandi (koefisien 987.650, p-value < 0.01), jumlah lantai (koefisien 542.411, p-value < 0.01), dan fasilitas parkir (koefisien 408.438, p-value 0.011) memiliki pengaruh signifikan terhadap harga. Sementara itu, faktor seperti akses jalan utama, kamar tamu, dan AC tidak menunjukkan pengaruh signifikan (p-value > 0.05). Model ini dapat digunakan untuk penilaian properti, perencanaan investasi, dan negosiasi harga.

ABSTRACT

The continuously increasing demand in Indonesia's property market affects house prices, which are influenced by various factors. Classification, as one of the methods in data mining, has the ability to make predictions. A house price prediction model was developed using linear regression techniques. Analysis results indicate that the main factors affecting house prices are land area, number of bathrooms, number of floors, and parking facilities. This model is accurate, particularly for low to mid-priced houses. Factors such as land area (coefficient 242.186, very small p-value), number of bathrooms (coefficient 987.650, p-value < 0.01), number of floors (coefficient 542.411, p-value < 0.01), and parking facilities (coefficient 408.438, p-value 0.011) have a significant impact on prices. Meanwhile, factors like main road access, guest rooms, and air conditioning do not show significant effects (p-value > 0.05). This model can be used for property valuation, investment planning, and price negotiation.

Penulis Korespondensi:

Siti Khoiriyah,
Teknologi Informasi,
Universitas Ibrahimy, Situbondo, Jawa Timur, Indonesia
Email: sitikhoiriyah@gmail.com

This is an open access article under the [CC BY-SA](#) license

1. PENDAHULUAN

Perkembangan pasar properti di Indonesia menunjukkan dimana suatu barang yang dijual sangat dipengaruhi oleh sejumlah faktor, termasuk lokasi, kondisi fisik, dan dinamika pasar [1]. Rumah merupakan kebutuhan paling dasar manusia untuk keamanan [2]. Sebagai unit sosial terkecil, rumah memiliki peran yang

sangat penting dalam keberadaan manusia [3]. Tidak hanya sebagai tempat tinggal, rumah adalah pusat untuk segala jenis aktivitas, mulai dari pekerjaan rumah sehari-hari hingga refleksi momen-momen penting [4].

Istilah data mining pertama kali menjadi populer pada tahun 1990-an [5]. Data mining adalah proses menganalisis kumpulan data besar untuk mengungkapkan pola, tren, atau pengetahuan yang berguna [6]. Data memiliki potensi untuk mengungkap wawasan yang dapat membantu dalam membuat suatu keputusan [7]. Ada banyak algoritma dan teknik yang digunakan untuk menganalisis data dan mengungkap pola tersembunyi. Data mining juga merupakan proses yang menggunakan algoritma untuk mengekstrak pengetahuan implisit dari data [8]. Sebagai subdisiplin ilmu komputer, data mining merupakan model yang relatif baru yang mencakup pengumpulan, pembersihan, dan analisis data [9]. Klasifikasi merupakan bentuk data mining yang memiliki kemampuan melakukan prediksi [10]. Tujuan utama data mining adalah untuk mengolah sebuah data yang dapat memberikan informasi baru [11].

Kompleksitas informasi harga rumah membuat konsumen kesulitan membuat keputusan pembelian. Penentuan harga yang akurat sangat penting dalam transaksi properti. Untuk mengatasi masalah tersebut, perlu dikembangkan model simulasi prediksi harga rumah yang mempertimbangkan berbagai variabel, termasuk luas tanah, luas bangunan, dan lokasi. Model algoritma Linear Regression akan digunakan dalam penelitian ini untuk memprediksi harga rumah. Algoritma Linear Regression adalah metode yang digunakan untuk mengidentifikasi hubungan antara satu varabel dependen dengan satu atau lebih variabel idenpenden. Pemilihan algoritma ini dipilih karena memiliki kemampuan untuk melakukan perhitungan secara efisien. Dataset yang digunakan dalam penelitian ini adalah dataset publik dari Kaggle.

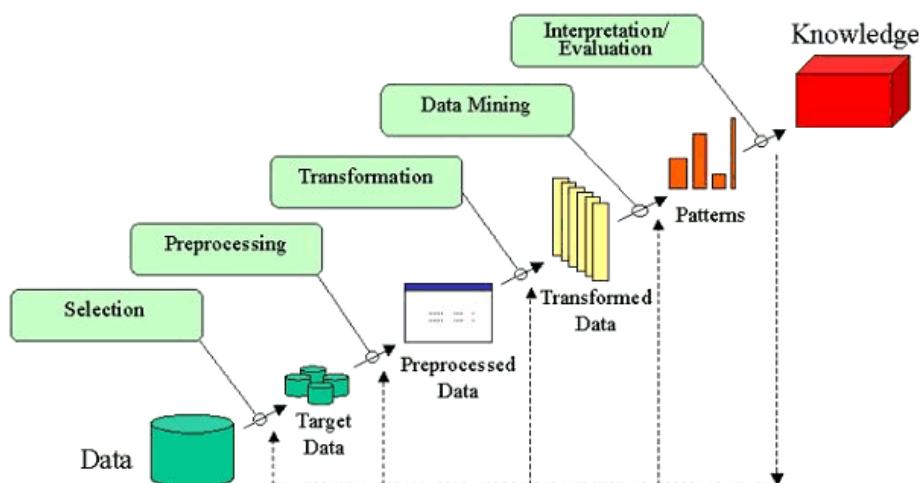
2. METODE PENELITIAN

2.1. Pengumpulan Data

Situs Kaggle menyediakan dataset yang digunakan dalam penelitian ini. Dengan memanfaatkan koleksi harga rumah yang dikenal dengan "*Housing Price Dataset*" yang disusun oleh M Yasser H. Dataset berisi sebanyak 13 atribut yaitu harga rumah, luas bangunan, jumlah kamar tidur, jumlah kamar mandi, jumlah lantai, jalan utama, kamar tamu, ruang bawah tanah, pemanas air, AC, tempat parkir, prefarea dan furnitur dengan sampel sebanyak 545 data.

2.2. Proses Desain

Proses desain untuk menerapkan data mining dalam penelitian ini menggunakan Knowledge Discovery in Databases (KDD) dan memanfaatkan perangkat lunak RapidMiner. Ilustrasi visual dari tahap-tahap KDD terdapat pada Gambar 1.



Gambar 1 . Diagram Konseptual Model KDD

Proses KDD terdiri dari lima fase sebagai berikut:

- **Pemilihan data (Selection data):** Data yang relevan akan dikumpulkan dan dipersiapkan pada tahap ini. Atribut dari dataset real estat akan dianalisis berdasarkan kebutuhan penelitian. Ada sebanyak 13 atribut dalam dataset yang akan digunakan.

- Mempersiapkan data (*Pre-processing data*): Pembersihan data untuk data duplikat, nilai yang hilang, data yang identik, atau data yang tidak berkorelasi dengan variabel lain.
- Transformasi data (*Data transformation*): Mentransformasi data ke format yang sesuai untuk jenis data ke dalam format yang sesuai untuk diproses data mining. Proses transformasi melibatkan modifikasi peran atribut yang akan digunakan sebagai identitas dan label.
- Data mining: Langkah ini merupakan metode utama untuk mengidentifikasi data atau informasi yang telah dikumpulkan melalui proses sebelumnya dengan menerapkan metode yang sesuai. Dalam penelitian ini, algoritma Linear Regression digunakan untuk memprediksi harga rumah. Rumus yang digunakan untuk Linear Regression sebagai berikut:

$$Y_i = f(X_i, \beta) + \varepsilon_i \quad (1)$$

Variabel respons, yang juga disebut sebagai variabel dependen, adalah Y_i . X_i adalah prediktor atau variabel independen. Nilai variabel dependen diprediksi menggunakan variabel ini. Variabel independen (X_i) dan variabel dependen (Y_i) terhubung oleh fungsi f . Fungsi ini sering kali berbentuk garis lurus dalam regresi linier. Parameter yang tidak diketahui dari model regresi dilambangkan dengan β . Data yang tersedia akan digunakan untuk memperkirakan parameter ini. Kemiringan dan titik potong garis regresi ditentukan oleh parameter ini. Istilah " ε_i " merujuk pada kesalahan atau residu. Frasa ini menunjukkan perbedaan antara nilai aktual variabel dependen (Y_i) dan nilai yang diprediksi oleh model regresi. Menurut rumus sebelumnya, sebuah fungsi (f) dari variabel independen (X_i) ditambah kesalahan (ε_i) dapat digunakan untuk memprediksi atau menjelaskan nilai variabel dependen (Y_i). Karena regresi linier sederhana biasanya menggambarkan fungsi f ini sebagai garis lurus, persamaan di atas dapat dinyatakan sebagai berikut:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \quad (2)$$

dimana :

β_0 : Titik potong garis regresi dengan sumbu Y.

β_1 : Kemiringan garis regresi.

- Interpretasi dan evaluasi data: Langkah terakhir dari prosedur ini melibatkan pola-pola yang dihasilkan oleh data mining.

3. HASIL DAN ANALISIS

3.1. Hasil Koleksi Data

Data yang digunakan dalam penelitian ini diperoleh dari situs web Kaggle.com. Dalam data ini, terdapat 13 atribut: usia, harga, area, kamar tidur, kamar mandi, jumlah lantai, jalan utama, ruang tamu, ruang bawah tanah, pemanas air, pendingin udara, parkir, area yang diinginkan, dan status furnitur. Jumlah data yang diperoleh adalah 545 sampel.

price	area	bedrooms	bathroom	stories	mainroad	guestroom	basement	hotwater	airconditi	parking	prefarea	furnishingstatus
13300000	7420	4	2	3	yes	no	no	yes	2	yes	furnished	
12250000	8960	4	4	4	yes	no	no	yes	3	no	furnished	
12250000	9960	3	2	2	yes	no	yes	no	2	yes	semi-furnished	
12215000	7500	4	2	2	yes	no	yes	no	yes	3	yes	furnished
...
1767150	2400	3	1	1	no	no	no	no	0	no	semi-furnished	
1750000	3620	2	1	1	yes	no	no	no	0	no	unfurnished	
1750000	2910	3	1	1	no	no	no	no	0	no	furnished	
1750000	3850	3	1	2	yes	no	no	no	0	no	unfurnished	

Gambar 2. Harga Rumah

3.2. Hasil Pemilihan Data

Setiap atribut dalam dataset harga rumah ini akan dipilih dan di proses pada langkah berikutnya. Pendekatan komprehensif ini memastikan bahwa tidak ada informasi relevan yang terlewatkan, memaksimalkan potensi untuk analisis yang akurat dan mendalam.

3.3. Hasil Mempersiapkan Data

Proses dalam data mining yang disebut pemrosesan data melibatkan mengubah data yang belum diproses menjadi informasi yang berguna. Pada tahapan ini, dilakukan pembersihan data dari anomali atau nilai-nilai anomali yang mungkin mengganggu proses analisis yang disebut dengan proses outlier.

price	outlier	area	bedrooms	bathrooms	stories	mainroad	guestroom	basement	hotwaterhe...	airconditioni...	parking
13300000	false	7420	4	2	3	yes	no	no	no	yes	2
12250000	false	8960	4	4	4	yes	no	no	no	yes	3
12250000	false	9960	3	2	2	yes	no	yes	no	no	2
12215000	false	7500	4	2	2	yes	no	yes	no	yes	3
11410000	false	7420	4	1	2	yes	yes	yes	no	yes	2
10850000	false	7500	3	3	1	yes	no	yes	no	yes	2
10150000	false	8580	4	3	4	yes	no	no	no	yes	2
10150000	true	16200	5	3	2	yes	no	no	no	no	0
9870000	false	8100	4	1	2	yes	yes	yes	no	yes	2
9800000	false	5750	3	2	4	yes	yes	no	no	yes	1
9800000	true	13200	3	1	2	yes	no	yes	no	yes	2
9681000	false	6000	4	3	2	yes	yes	yes	yes	no	2
9310000	false	6550	4	2	2	yes	no	no	no	yes	1
9240000	false	3500	4	2	2	yes	no	no	yes	no	2
9240000	false	7800	3	2	2	yes	no	no	no	no	0
9100000	false	6000	4	1	2	yes	no	yes	no	no	2
9100000	false	6600	4	2	2	yes	yes	yes	no	yes	1
8950000	false	8500	3	2	4	yes	no	no	no	yes	2

Gambar 3. Hasil dari Proses Outlier

Pada gambar 3 merupakan outlier dengan titik data diklasifikasikan sebagai true atau false. Nilai salah menunjukkan bahwa data tersebut tidak dianggap sebagai pencilan, sedangkan nilai benar menunjukkan sebagai pencilan. Selain itu, terdapat beberapa data yang memiliki nilai nol. Langkah selanjutnya adalah memfilter data yang miliki nilai true dan nol. Hasil proses filter ditunjukkan pada gambar 4.

price	outlier	area	bedrooms	bathrooms	stories	mainroad	guestroom	basement	hotwaterhe...	airconditioni...	parking
13300000	false	7420	4	2	3	yes	no	no	no	yes	2
12250000	false	8960	4	4	4	yes	no	no	no	yes	3
12250000	false	9960	3	2	2	yes	no	yes	no	no	2
12215000	false	7500	4	2	2	yes	no	yes	no	yes	3
11410000	false	7420	4	1	2	yes	yes	yes	no	yes	2
10850000	false	7500	3	3	1	yes	no	yes	no	yes	2
10150000	false	8580	4	3	4	yes	no	no	no	yes	2
9870000	false	8100	4	1	2	yes	yes	yes	no	yes	2
9800000	false	5750	3	2	4	yes	yes	no	no	yes	1
9681000	false	6000	4	3	2	yes	yes	yes	yes	no	2
9310000	false	6550	4	2	2	yes	no	no	no	yes	1
9240000	false	3500	4	2	2	yes	no	no	yes	no	2
9100000	false	6000	4	1	2	yes	no	yes	no	no	2
9100000	false	6600	4	2	2	yes	yes	yes	no	yes	1
8950000	false	8500	3	2	4	yes	no	no	no	yes	2
8890000	false	4600	3	2	2	yes	yes	no	no	yes	2
8855000	false	6420	3	2	2	yes	no	no	no	yes	1

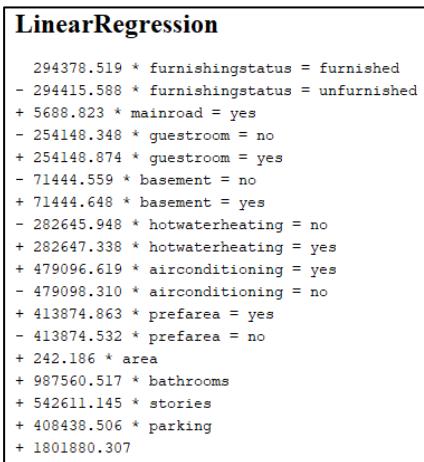
Gambar 4. Hasil dari Memfilter Data yang Bernilai True dan Nol

Attributes	area	bedrooms	bathrooms	stories	mainroad	guestroom	basement	hotwaterhe...	airconditio...	parking	prefarea	furnishing...
area	1	0.130	0.161	0.079	-0.253	0.140	0.073	-0.105	-0.251	0.178	-0.316	-0.183
bedrooms	0.130	1	0.388	0.435	0.040	0.115	0.083	0.025	-0.183	0.113	-0.071	-0.139
bathrooms	0.161	0.388	1	0.393	-0.039	0.120	0.096	0.053	-0.205	0.071	0.002	-0.157
stories	0.079	0.435	0.393	1	-0.090	0.040	-0.271	0.038	-0.278	-0.023	0.068	-0.072
mainroad	-0.253	0.040	-0.039	-0.090	1	-0.037	-0.029	-0.017	0.110	-0.238	0.174	0.129
guestroom	0.140	0.115	0.120	0.040	-0.037	1	0.359	-0.055	-0.154	-0.045	-0.165	-0.132
basement	0.073	0.083	0.096	-0.271	-0.029	0.359	1	-0.063	-0.067	0.100	-0.207	-0.152
hotwaterhe...	-0.105	0.025	0.053	0.038	-0.017	-0.055	-0.063	1	0.216	-0.043	0.082	-0.043
aircondition...	-0.251	-0.183	-0.205	-0.278	0.110	-0.154	-0.067	0.216	1	-0.088	0.071	0.225
parking	0.178	0.113	0.071	-0.023	-0.238	-0.045	0.100	-0.043	-0.088	1	-0.082	-0.052
prefarea	-0.316	-0.071	0.002	0.068	0.174	-0.165	-0.207	0.082	0.071	-0.082	1	0.023
furnishing...	-0.183	-0.139	-0.157	-0.072	0.129	-0.132	-0.152	-0.043	0.225	-0.052	0.023	1

Gambar 5. Hasil dari Korelasi Matriks

Evaluasi korelasi matriks pada gambar 5 dilakukan setelah memfilter data. Korelasi matriks menggambarkan beberapa titik properti dataset saling berhubungan satu sama lain. Matriks ini memiliki angka yang berkisar dari -1 hingga 1, di mana 1 menunjukkan korelasi positif yang sempurna, yang berarti bahwa ketika satu atribut meningkat, atribut lainnya cenderung meningkat juga. Di sisi lain, korelasi negatif sempurna diwakili oleh nilai -1, yang berarti bahwa ketika satu atribut meningkat, atribut lainnya cenderung menurun. Terakhir, 0 menunjukkan tidak ada korelasi, yang berarti tidak ada hubungan yang berarti antara kedua atribut tersebut. Dapat disimpulkan bahwa terdapat hubungan positif kuat antara jumlah kamar mandi dan kamar tidur. Selain itu, jumlah kamar dan ukuran area memiliki hubungan positif, menunjukkan bahwa rumah yang lebih besar biasanya memiliki lebih banyak kamar. Keberadaan jalan utama berkorelasi negatif dengan ukuran area, hal ini menunjukkan bahwa rumah besar lebih sering ditemukan di lingkungan yang lebih tenang. Selain itu, ada asosiasi negatif antara jumlah kamar tidur dan pendingin udara, yang menunjukkan hubungan negatif antara keduanya dan banyak pasangan atribut yang mendekati 0 memiliki korelasi yang rendah. Ini menunjukkan bahwa kedua sifat tersebut tidak secara signifikan saling terkait satu sama lain.

Untuk menerapkan regresi linier, matriks korelasi harus diikuti dengan konversi nilai nominal menjadi nilai numerik. *Output* algoritma menunjukkan bahwa koefisien area tersebut adalah 242.186. Ini menyiratkan bahwa harga properti akan meningkat sebesar 242.186 unit (dalam rupiah) untuk setiap peningkatan satuan persegi. Hal ini ditunjukkan pada gambar 6.



```
LinearRegression

294378.519 * furnishingstatus = furnished
- 294415.588 * furnishingstatus = unfurnished
+ 5688.823 * mainroad = yes
- 254148.348 * guestroom = no
+ 254148.874 * guestroom = yes
- 71444.559 * basement = no
+ 71444.648 * basement = yes
- 282645.948 * hotwaterheating = no
+ 282647.338 * hotwaterheating = yes
+ 479096.619 * airconditioning = yes
- 479098.310 * airconditioning = no
+ 413874.863 * prefarea = yes
- 413874.532 * prefarea = no
+ 242.186 * area
+ 987560.517 * bathrooms
+ 542611.145 * stories
+ 408438.506 * parking
+ 1801880.307
```

Gambar 6. Hasil dari Linear Regression

3.4. Transformasi Data

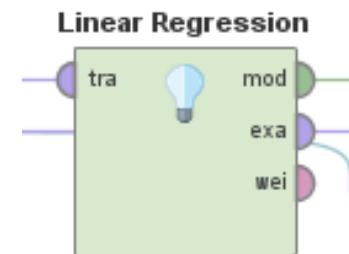
Pada tahap ini memodifikasi struktur, format, atau nilai dataset agar memenuhi persyaratan analisis atau pemodelan data. Tujuan pada tahap ini adalah untuk memodifikasi struktur, format, atau nilai dataset agar memenuhi persyaratan analisis atau pemodelan data yang akan datang. Metode normalisasi transformasi rentang min-max ditentukan oleh parameter. Ini melibatkan melakukan terjemahan data linier dengan menggunakan nilai minimum dan maksimum. Rentang nilai untuk parameter ini adalah 0 hingga 1, di mana 0 adalah nilai terendah dan 1 adalah nilai tertinggi.

price	area	bedrooms	bathrooms	stories	parking	mainroad = ...	mainroad = ...	guestroom ...	guestroom ...	basement = ...	basement
13300000	0.397	0.600	0.333	0.667	0.667	1	0	1	0	1	0
12250000	0.502	0.600	1	1	1	1	0	1	0	1	0
12250000	0.571	0.400	0.333	0.333	0.667	1	0	1	0	0	1
12215000	0.402	0.600	0.333	0.333	1	1	0	1	0	0	1
11410000	0.397	0.600	0	0.333	0.667	1	0	0	1	0	1
10850000	0.402	0.400	0.667	0	0.667	1	0	1	0	0	1
10150000	0.476	0.600	0.667	1	0.667	1	0	1	0	1	0
10150000	1	0.800	0.667	0.333	0	1	0	1	0	1	0
9870000	0.443	0.600	0	0.333	0.667	1	0	0	1	0	1
9800000	0.282	0.400	0.333	1	0.333	1	0	0	1	1	0
9800000	0.794	0.400	0	0.333	0.667	1	0	1	0	0	1
9681000	0.299	0.600	0.667	0.333	0.667	1	0	0	1	0	1
9310000	0.337	0.600	0.333	0.333	0.333	1	0	1	0	1	0
9240000	0.127	0.600	0.333	0.333	0.667	1	0	1	0	1	0
9240000	0.423	0.400	0.333	0.333	0	1	0	1	0	1	0
9100000	0.299	0.600	0	0.333	0.667	1	0	1	0	0	1
9100000	0.340	0.600	0.333	0.333	0.333	1	0	0	1	0	1
9060000	0.471	0.400	0.333	1	0.667	1	0	1	0	1	0

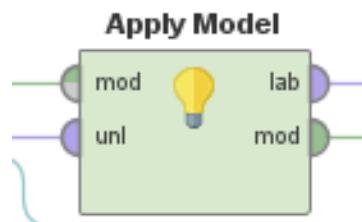
Gambar 7. Hasil dari Transformasi Data

3.5. Data Mining

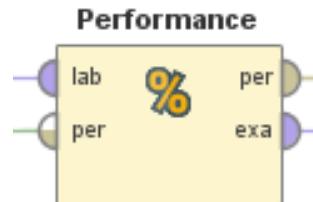
Pada tahap ini adalah fase utama dalam membuat model prediktif atau fase implementasi metode. Untuk menentukan hubungan antara satu variabel target (variabel dependen) dan satu atau lebih faktor prediktor (variabel independen), model regresi linier digunakan.

**Gambar 8. Algoritma Linear Regression**

Model ini diberi data pelatihan, yang terdiri dari pasangan nilai antara variabel prediktor dan variabel target untuk menentukan garis lurus mana yang paling baik menggambarkan hubungan antara faktor-faktor ini.

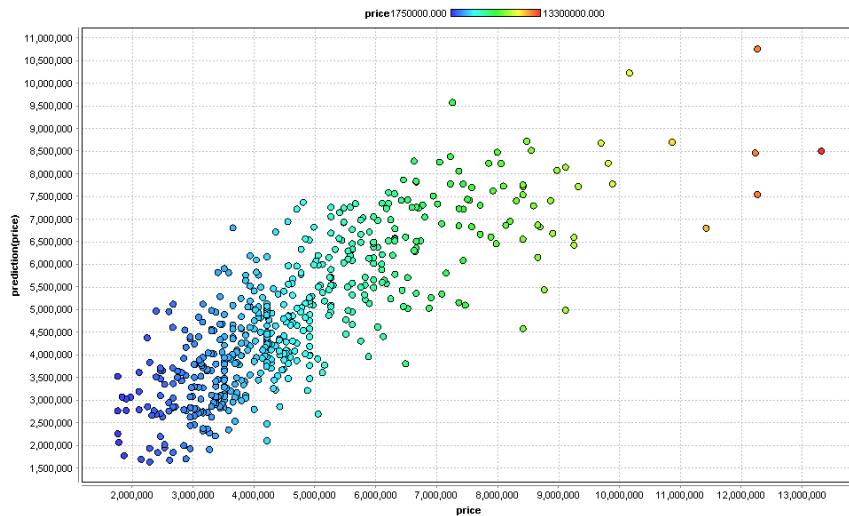
**Gambar 9. Apply Model**

Apply Model akan diterapkan pada data baru setelah dilatih. Apply Model akan meramalkan nilai yang diinginkan dari data baru menggunakan persamaan yang telah dihasilkan.

**Gambar 10. Performance**

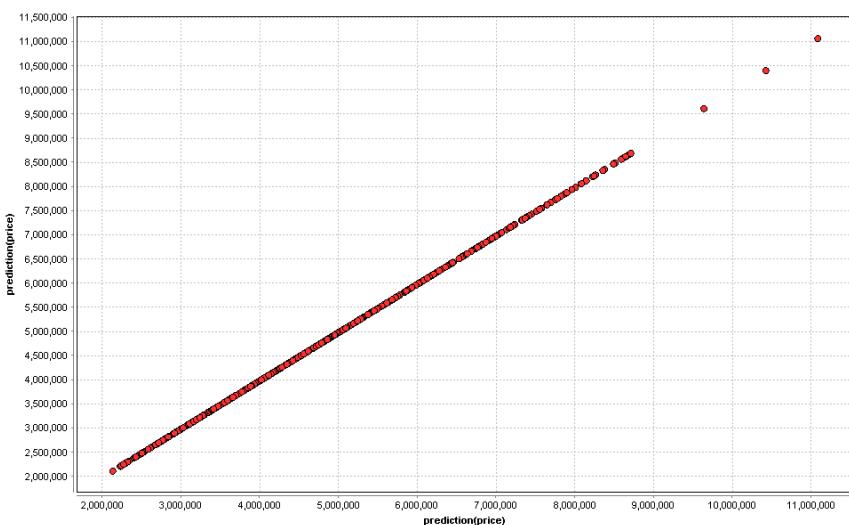
Tahap terakhir adalah evaluasi kinerja model. (Performance). Hasil peramalan model akan dibandingkan dengan nilai target aktual dari data terbaru. Perbandingan ini digunakan untuk membangun metrik kinerja model.

3.6. Interpretasi dan Evaluasi Data



Gambar 11. Grafik Perbandingan Harga Aktual dan Harga Prediksi

Tahap ini merupakan tahap evaluasi untuk memberikan wawasan dan mengukur efektivitas model algoritma Linear Regression. Harga sebenarnya dari properti tersebut diwakili oleh lokasi titik pada sumbu X, sementara harga yang diperkirakan oleh model ditunjukkan oleh lokasi titik pada sumbu Y. Setiap warna titik mewakili perbedaan antara harga aktual dan harga yang diantisipasi. Titik merah menunjukkan ketidaksesuaian yang lebih besar, tetapi titik biru biasanya memiliki biaya yang diperkirakan mendekati harga aktual.



Gambar 12. Grafik Harga Prediksi

Pada gambar 12 menunjukkan bahwa Algoritma Linear Regression dapat memperkirakan harga dengan akurasi tinggi karena terdapat titik-titik yang hampir membentuk garis lurus, hal ini menunjukkan bahwa harga yang diharapkan dan harga aktual cukup dekat. Distribusi titik yang padat menunjukkan tingkat kesalahan prediksi yang rendah. Faktor-Faktor penting variabel seperti luas, jumlah kamar mandi, jumlah lantai, dan parkir yang memiliki nilai p yang sangat rendah menunjukkan bahwa mereka secara signifikan mempengaruhi nilai properti.

Attribute	Coefficient	Std. Error	Std. Coefficient	Tolerance	t-Stat	p-Value	Code
furnishingstatus = unf...	-294415.588	293780.241	-0.071	0.817	-1.002	0.318	
mainroad = yes	5688.823	397891.448	0.001	0.914	0.014	0.989	
guestroom = no	-254148.348	379043.409	-0.045	0.889	-0.670	0.503	
guestroom = yes	254148.874	379043.409	0.045	0.889	0.671	0.503	
basement = no	-71444.559	270267.643	-0.017	0.959	-0.264	0.792	
basement = yes	71444.648	270267.643	0.017	0.959	0.264	0.792	
hotwaterheating = no	-282645.948	554453.288	-0.033	0.991	-0.510	0.611	
hotwaterheating = yes	282647.338	554453.288	0.033	0.991	0.510	0.611	
airconditioning = yes	479096.619	341005.221	0.112	0.642	1.405	0.162	
airconditioning = no	-479098.310	341005.221	-0.112	0.642	-1.405	0.162	
prefarea = yes	413874.863	321489.490	0.091	0.806	1.287	0.200	
prefarea = no	-413874.532	321489.490	-0.091	0.806	-1.287	0.200	
area	242.186	58.411	0.285	0.860	4.146	0.000	***
bathrooms	987560.517	273983.529	0.250	0.846	3.604	0.000	***
stories	542611.145	163537.605	0.221	0.918	3.318	0.001	**
parking	408438.506	158722.759	0.176	0.867	2.573	0.011	**

Gambar 12. Hasil Linear Regression

4. KESIMPULAN

Dari hasil analisis dapat disimpulkan bahwa model prediksi harga properti ini memiliki akurasi yang tinggi, dengan prediksi harga sesuai dengan harga aktual, terutama pada rentang harga rendah hingga menengah. Analisis regresi menunjukkan bahwa beberapa faktor berpengaruh signifikan terhadap harga properti, dilihat dari p-value yang rendah. Luas area memiliki koefisien 242.186 dengan p-value sangat kecil, sehingga setiap tambahan satu unit area dapat meningkatkan harga sekitar 242.186. Jumlah kamar mandi memiliki koefisien 987.650 dengan p-value < 0.01, artinya setiap tambahan satu kamar mandi dapat menaikkan harga sekitar 987.650. Jumlah lantai juga signifikan, dengan koefisien 542.411 dan p-value < 0.01, di mana setiap lantai tambahan menambah harga sekitar 542.411. Ketersediaan parkir memiliki koefisien 408.438 dan p-value 0.011, sehingga setiap ruang parkir tambahan menaikkan harga sekitar 408.438. Sebaliknya, faktor seperti akses jalan utama, kamar tamu, dan AC memiliki p-value di atas 0.05, yang menunjukkan bahwa pengaruhnya tidak signifikan terhadap harga.

REFERENSI

- [1] J. Septriaznu, "Rancang Bangun Sistem Prediksi Nilai Jual Properti Berbasis Web Menggunakan Regresi Linear," *Jurnal Sains dan Teknologi (JSIT)*, vol. 4, no. 1, pp. 27–45, Jan. 2024, doi: 10.47233/jsit.v4i1.1464.
- [2] I. Algoritma *et al.*, "Implementasi Algoritma Support Vector Regression dan Linear Regression Untuk Prediksi Harga Rumah".
- [3] A. Widyastuti, "Terbit online pada laman web jurnal: <http://ejurnal.unim.ac.id/index.php/submit> SUBMIT (Jurnal Ilmiah Teknologi Informasi dan Sains) PREDIKSI HARGA RUMAH SESUAI SPESIFIKASI MENGGUNAKAN METODE MULTIPLE LINEAR REGRESSION PREDICTION OF HOUSE PRICES ACCORDING TO SPECIFICATIONS USING MULTIPLE LINEAR REGRESSION METHOD," vol. 4, no. 1, pp. 30–35, 2024, [Online]. Available: <http://ejurnal.unim.ac.id/index.php/submit>
- [4] N. Khalijah Zainal, "PREDIKSI HARGA REAL ESTATE MENGGUNAKAN METODE REGRESI LINEAR BERBASIS MACHINE LEARNING," 2024.
- [5] J. Han, J. Pei, and H. Tong, *Data Mining: Concepts and Techniques*. in The Morgan Kaufmann Series in Data Management Systems. Morgan Kaufmann, 2022. [Online]. Available: <https://books.google.co.id/books?id=NR1oEAAAQBAJ>
- [6] S. K. M. T. M. M. I. P. M. Dr. Ir. N. Tri S. Saptadi *et al.*, *DATA MINING*. 2024. [Online]. Available: <https://books.google.co.id/books?id=CR4mEQAAQBAJ>
- [7] S. K. M. M. S. I. Ir. T. Irfan Fajri *et al.*, *Data Mining*. Serasi Media Teknologi, 2024. [Online]. Available: <https://books.google.co.id/books?id=YykdeEQAAQBAJ>

-
- [8] S. T. M. K. Yahya, *Data Mining*. CV Jejak (Jejak Publisher), 2022. [Online]. Available: <https://books.google.co.id/books?id=0J2mEAAAQBAJ>
 - [9] S. K. M. K. D. A. N. A. P. S. K. M. K. Amril Mutoi Siregar, *DATA MINING: Pengolahan Data Menjadi Informasi dengan RapidMiner*. CV Kekata Group. [Online]. Available: <https://books.google.co.id/books?id=rTImDwAAQBAJ>
 - [10] S. Abrori and Z. Fatah, "Implementasi Metode CNN Untuk Klasifikasi Status Stunting Pada Balita," *Gudang Jurnal Multidisiplin Ilmu*, vol. 2, no. 10, pp. 380–385, 2024.
 - [11] F. A. Sianturi, P. M. Hasugian, A. Simangunsong, B. Nadeak, and H. T. Sihotang, *DATA MINING: Teori dan Aplikasi Weka*. in Edisi. IOCS Publisher, 2019. [Online]. Available: <https://books.google.co.id/books?id=MWcHEAAAQBAJ>