

PREDIKSI HARGA RUMAH MENGGUNAKAN LINEAR REGRESSION PADA PERUMAHAN KECAMATAN PAGUYANGAN

Matien Hakim Falahudin Bachtiar¹, Tezhar Rayendra Trastaronny Pastika Nugraha², Sorikhi³

^{1,2,3}Informatika, Fakultas Sains dan Teknologi, Universitas Peradaban.

matienkasa@gmail.com, tezhar.rayendra19@gmail.com, soryc2001@yahoo.com

Jl. Raya Pagojengan Km 3 Paguyangan, Brebes, Jawa Tengah, 52276

Keywords:

prediction, house
price prediction,
linear regression,
multiple linear
regression, z-score

Abstract

House prices in Indonesia fluctuate in line with economic growth. In Paguyangan District, factors such as land area, building area, number of rooms, electrical capacity, and proximity to main roads are key determinants of house prices. This study aims to assess the accuracy of a Linear Regression model in predicting house prices using 341 data points obtained through observation and interviews. The research stages include data pre-processing (data cleaning, categorical feature transformation, and outlier removal using the Z-Score method), data splitting using the 10-Fold Cross Validation technique, model training, and evaluation using R^2 , MSE, RMSE, and MAE metrics. The best result was achieved in the sixth fold with an R^2 of 0.9427, RMSE of IDR 13,320,382.01, and MAE of IDR 10,541,293.61, indicating a relatively small prediction error compared to actual prices. This value indicates that 94.27% of the variation in house prices can be explained by the developed regression model. These findings conclude that the Linear Regression model in this study demonstrates excellent performance and is suitable for application in the context of house price prediction.

Kata Kunci:

Prediksi, prediksi
harga rumah,
linear regression.
regresi linear
berganda, z-score

Abstrak

Harga rumah di Indonesia mengalami fluktuasi seiring pertumbuhan ekonomi. Di Kecamatan Paguyangan, faktor luas tanah, luas bangunan, jumlah kamar, daya listrik, dan kedekatan dengan jalan raya menjadi penentu harga rumah. Penelitian ini bertujuan mengukur akurasi model *Linear Regression* dalam memprediksi harga rumah menggunakan 341 data hasil observasi dan wawancara. Tahapan penelitian meliputi pre-processing (pembersihan data, transformasi fitur kategorikal, penghapusan outlier dengan metode *Z-Score*), pembagian data dengan teknik *10-Fold Cross Validation*, pelatihan model, serta evaluasi menggunakan R^2 , MSE, RMSE, dan MAE. Hasil terbaik pada *fold* ke-6 menunjukkan R^2 sebesar 0,9427, RMSE Rp 13.320.382,01, dan MAE Rp 10.541.293,61 yang menandakan tingkat kesalahan prediksi relatif kecil terhadap harga aktual. Nilai ini menunjukkan bahwa sebesar 94,27% variasi harga rumah dapat dijelaskan oleh model regresi yang dibangun. Hasil ini disimpulkan bahwa performa model *Linear Regression* dalam penelitian ini berada pada kategori sangat baik dan layak digunakan untuk konteks prediksi harga rumah.

I. Pendahuluan

Perkembangan sektor properti di Indonesia menunjukkan tren peningkatan yang cukup pesat dalam beberapa tahun terakhir. Salah satu aspek krusial dalam sektor ini adalah harga rumah, yang menjadi faktor utama dalam pengambilan keputusan oleh para calon pembeli [1]. Rumah merupakan tempat tinggal, tempat berkumpul, tempat berlindung yang ditempati oleh manusia

untuk melakukan aktivitas sehari-hari [2]. Hal ini menjadi suatu kebutuhan yang akan terus berkembang sesuai dengan kebutuhan dan tingkat pertumbuhan masyarakat yang semakin meningkat. Kecamatan Paguyangan sebagai salah satu wilayah dengan kawasan yang sedang mengalami perkembangan khususnya dalam sektor perumahan. Perumahan merupakan sekumpulan rumah yang digunakan untuk tempat tinggal yang dirancang dengan mempertimbangkan beberapa aspek seperti kualitas lingkungan yang baik, aman, tingkat keterjangkauan serta menciptakan kawasan yang asri dan nyaman [3].

Perkembangan perumahan yang ada di kawasan Kecamatan Paguyangan ini juga tentunya banyak menghadapi beberapa tantangan yang sangat bervariasi terutama dalam menentukan harga rumah yang dapat dipengaruhi oleh beberapa faktor diantaranya oleh lokasi, spesifikasi, dan fasilitas yang ada. Hal tersebut dapat menyebabkan kesulitan bagi calon pembeli dalam menentukan rumah sesuai dengan kebutuhan dan anggaran yang mereka miliki serta kurangnya informasi dan promosi dari agen properti menjadi hambatan bagi pembeli dalam mencari informasi mengenai harga rumah.

Faktor yang menjadi pengaruh dalam menentukan harga rumah ini sangat perlu diperhatikan sehingga perlu adanya pendekatan untuk memahami faktor yang menjadi pengaruh harga rumah secara akurat yang dapat memberikan solusi untuk calon pembeli dan agen properti dalam menghadapi tantangan tersebut. Oleh karena itu perlu dilakukannya prediksi harga rumah berdasarkan faktor-faktor yang menjadi pengaruhnya.

Prediksi merupakan metode yang digunakan untuk memperkirakan sesuatu dengan melalui tahap pertimbangan dari data atau informasi yang didapat sebelumnya, hasil prediksi tidak selalu menjadi hal yang akan terjadi tetapi sesuatu yang ada yang hampir mendekati dengan realitasnya [4]. Metode yang digunakan dalam memprediksi harga rumah sangat bervariasi juga, diantaranya seperti *Linear Regression*, *Random Forest*, dan *K-Nearest Neighbor*, metode tersebut biasa digunakan untuk mengolah data dan menghasilkan hasil prediksi yang akurat terutama dalam memprediksi harga rumah.

Setiap metode yang sebelumnya telah digunakan dalam memprediksi harga rumah memiliki kelebihan dan kelemahan masing-masing sehingga diperlukan pemahaman mengenai karakteristik masing-masing metode yang ada untuk menentukan metode yang cocok untuk situasi yang terjadi pada penelitian kali ini. Salah satu metode yang digunakan untuk memprediksi harga rumah yaitu metode *Linear Regression* yang merupakan suatu metode yang digunakan untuk melihat, memodelkan, hubungan yang terjadi antara dua atau lebih variabel dependen (harga rumah) dengan lebih variabel (spesifikasi yang mempengaruhi harga rumah) [1].

Peneliti kali ini dapat mempertimbangkan berbagai metode dari masing-masing kelebihan dan kekurangannya, seperti dari hal akurasi, interpretabilitas serta efisiensi komputasinya. Metode yang telah dijelaskan sebelumnya seperti *Random Forest* dan *K-Nearest Neighbor* menunjukkan akurasi yang tinggi, metode tersebut memiliki model yang lebih kompleks sehingga waktu prediksi memerlukan waktu lebih lama, terutama pada data yang besar. Selain itu, metode *Decision Tree* yang mudah dipahami namun sangat rentan terhadap *overfitting* dan perubahan data baru sehingga peneliti mempertimbangkan metode yang dipilih dalam penelitian kali ini dalam melakukan prediksi rumah yaitu menggunakan metode *Linear Regression*, karena memiliki karakteristik yang sederhana serta lebih mudah diimplementasikan dan efektif digunakan dalam dataset berskala kecil seperti di kawasan Kecamatan Paguyangan ini.

Metode *Linear Regression* ini juga cocok digunakan untuk hubungan linear antar variabel variabel meski metode ini memiliki keterbatasan dalam menangkap pola *non-linear* yang kompleks sehingga penelitian kali ini akan meninjau sejauh mana metode *Linear Regression* ini dapat memberikan hasil prediksi yang lebih akurat dalam memprediksi harga rumah

berdasarkan faktor-faktor yang mempengaruhinya di wilayah Kecamatan Paguyangan ini. Keterbatasan yang dimiliki oleh metode *Linear Regression* ini dapat diperbaiki dengan melakukan kombinasi antara pendekatan Regresi Linear dengan teknik pemrosesan data lanjutan dengan memperkuat variabel input dan transformasi *non-linear* terhadap data yang memiliki hubungan tidak linear terhadap harga. Penerapan validasi silang (*K-Fold Cross Validation*) juga menjadi langkah penting guna memastikan bahwa model regresi yang dibangun memiliki kemampuan generalisasi yang baik terhadap data baru, terutama dalam konteks harga rumah di kawasan yang sedang berkembang seperti Kecamatan Paguyangan.

Tujuan dari penelitian ini adalah untuk membangun dan mengevaluasi model prediksi harga rumah di Kecamatan Paguyangan menggunakan metode *Linear Regression* dengan meninjau efektivitasnya dalam mengolah faktor-faktor yang memengaruhi harga rumah serta menerapkan strategi perbaikan model melalui teknik pemrosesan data lanjutan dan validasi silang guna meningkatkan akurasi dan kemampuan generalisasi model terhadap data yang dinamis sehingga diharapkan dapat memberikan hasil yang praktis serta akurat dalam menentukan harga rumah dan informasi sesuai dengan kebutuhan yang diinginkan.

II. Landasan Teori

1. Prediksi

Menurut Kamus Besar Bahasa Indonesia (KBBI) prediksi adalah memperkirakan, atau ramalan. Prediksi merupakan salah satu metode memperkirakan sesuatu yang dapat menghasilkan nilai atau informasi pada masa yang akan datang dengan mempertimbangkan segala aspek berupa data maupun informasi dari masa lampau maupun masa sekarang [4].

2. Harga Rumah

Harga rumah merupakan penetapan suatu nilai atau jumlah uang yang harus diperhitungkan dari penjual kepada pembeli untuk dibayarkan apabila sudah mencapai kesepakatan. Penetapan harga rumah biasanya dapat diperhitungkan dari berbagai aspek seperti berdasarkan luas tanah, bangunan dan beberapa kriteria yang lain menjadi pokok dalam penjualan rumah [5].

3. Prediksi Harga Rumah

Prediksi harga rumah merupakan memperkirakan suatu nilai atau harga untuk rumah berdasarkan faktor yang mempengaruhi seperti luas tanah, bangunan, jumlah kamar tidur, kamar mandi dan lain-lain. Prediksi harga rumah digunakan sebagai upaya dalam membantu pembeli atau investor dalam pengambilan keputusan terkait proses pembelian rumah yang telah melalui berbagai aspek perbandingan sebelum keputusan yang ditentukan. Prediksi ini bertujuan untuk memberikan informasi yang akurat dan mendekati realitas agar dapat membantu pihak dalam membuat keputusan [1].

4. *Linear Regression*

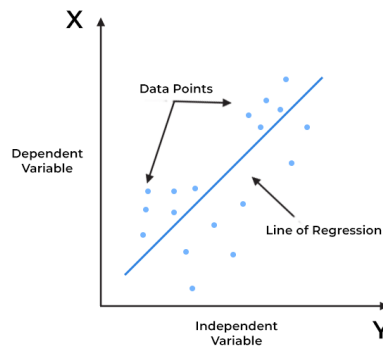
Linear Regression merupakan suatu metode statistik yang dapat memprediksi dengan menggunakan hubungan yang sistematis antara satu atau lebih variabel independen dengan satu variabel dependen [1].

Linear Regression ini lebih banyak diterapkan diberbagai bidang salah satunya dalam bidang ekonomi, lingkungan maupun ilmu sosial. Penelitian kali ini menggunakan metode *Linear Regression* untuk menentukan hasil dari hubungan variabel-variabel yang ada dalam masalah penelitian terkait prediksi harga rumah. Variabel-variabel yang digunakan dalam metode ini diharapkan nantinya dapat menghasilkan nilai akhir yang sesuai dengan metode yang digunakan, mengenai hasil prediksi dapat akurat sesuai dengan hasil aktual atau tidak sehingga dalam metode ini adanya perhitungan-perhitungan yang digunakan, sebagai berikut:

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p + \varepsilon$$

dengan:

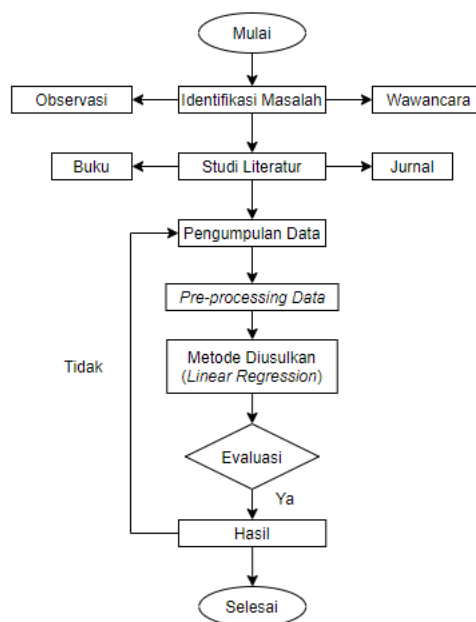
- Y = variabel dependen (variabel tak terikat).
- X₁ = Variabel dependen (variabel yang akan digunakan sebagai prediksi utama) Variabel independen (variabel yang mempengaruhi faktor prediksi).
-
- X_p
- β₀ = *Intercept* (nilai konstanta) adalah titik dimana garis regresi memotong sumbu y.
- β_{1, ..., β_p} = β_{1, β_{2, ..., β_p} adalah koefisien regresi yang menunjukkan rata-rata pada perubahan Y ketika X_{1, X_{2, ..., X_p} meningkat sebesar satu-satuan. Kesalahan atau residual adalah selisih antara nilai sebenarnya dengan nilai yang diprediksi oleh model regresi [1].}}



Gambar 1 Grafik Metode Linear Regression

III. Metodologi

1. Tahapan Penelitian



Gambar 2 Tahapan Penelitian

Tahapan-tahapan yang telah dirancang dalam penelitian, dapat dijelaskan mengenai penjelasan masing-masing dari tahapan penelitian diatas, sebagai berikut:

2. Identifikasi Masalah

Tahapan penelitian kali ini yang dilakukan peneliti pertama kali yaitu mengidentifikasi masalah yang terjadi dilingkungan sekitar yang akan menjadi objek penelitian. Berdasarkan hasil observasi di Perumahan Palm Indah Kecamatan Paguyangan ditemukan bahwa kawasan tersebut terdiri dari berbagai tipe unit hunian dengan kondisi fisik bangunan yang umumnya terawat serta didukung oleh infrastruktur lingkungan yang cukup memadai seperti akses jalan yang baik. Selain itu, perumahan ini juga masih dalam tahap pengembangan dengan pembangunan rumah-rumah baru yang siap dihuni.

3. Studi Literatur

Pengumpulan data dan informasi dalam penelitian ini tidak terlepas dari pemanfaatan berbagai referensi dan sumber-sumber ilmiah yang relevan. Referensi tersebut mencakup jurnal-jurnal penelitian yang membahas terkait prediksi harga rumah menggunakan *Linear Regression*, buku-buku akademik seperti buku *Hands-On Application Development with PyCharm*, serta literatur lainnya yang memiliki keterkaitan dengan permasalahan yang dikaji dalam penelitian ini. Penggunaan sumber-sumber tersebut bertujuan untuk memperkuat landasan teori, mendukung analisis dan memberikan perspektif yang lebih luas terhadap permasalahan yang diteliti.

4. Pengumpulan Data

Tahap pengumpulan data menjadi langkah awal yang penting dalam pengembangan sistem prediksi harga rumah menggunakan algoritma *Linear Regression*. Data diperoleh secara langsung sebagai data sekunder dari pihak pengelola Perumahan Palm Indah Kecamatan Paguyangan. Data tersebut berisi spesifikasi setiap unit rumah, antara lain luas tanah, luas bangunan, jumlah kamar tidur, jumlah kamar mandi, daya listrik, kedekatan dengan jalan utama, serta harga rumah.

5. Metode yang diusulkan

Metode yang diusulkan dalam penelitian ini adalah *Linear Regression* (Regresi Linier Berganda) yang digunakan untuk memodelkan hubungan antara satu variabel terikat (harga rumah) dan beberapa variabel bebas (faktor-faktor yang mempengaruhi harga rumah). Pendekatan ini dipilih karena *Linear Regression* memiliki karakteristik yang sederhana, mudah diinterpretasikan, dan efektif untuk digunakan pada dataset dengan skala kecil hingga menengah seperti di kawasan Perumahan Palm Indah, Kecamatan Paguyangan.

6. Evaluasi Data

Tahapan ini peneliti menguji hasil prediksi dari akurasi dengan menggunakan MSE (*Mean Squared Error*), RMSE (*Root Mean Squared Error*), MAE (*Mean Absolute Error*) dan *R-squared* (R^2).

Adapun beberapa penjelasan mengenai rumus dari perhitungan akurasi yang akan digunakan dalam penelitian ini yaitu:

$$RSME = \sqrt{\left(\frac{1}{n}\right) \sum (y_i - \hat{y})^2}$$

Perhitungan *R-squared* (R^2) pada persamaan prediksi yaitu sebagai berikut:

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

Perhitungan MAE pada persamaan prediksi yaitu sebagai berikut:

$$MAE = \frac{1}{N} \sum_{i=1}^N [x_i - y_i]$$

Perhitungan MSE pada persamaan prediksi yaitu sebagai berikut:

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y})^2$$

IV. Hasil dan Pembahasan

1. Pengumpulan Dataset

Data dikumpulkan secara langsung melalui observasi dan wawancara, yang kemudian disimpan dalam file Excel dengan format .xlsx. Dari hasil pengumpulan data melalui observasi dan wawancara didapatkan data sebanyak 341 data rumah yang ada pada perumahan Palm Indah Kecamatan Paguyangan, data-data tersebut mencakup informasi seperti nama konsumen, alamat rumah (blok dan nomor rumah), type dan yang terpenting spesifikasi yang mempengaruhi harga rumah seperti luas tanah (LT), luas bangunan (LB), jumlah kamar, daya listrik, dan kedekatan dengan jalan utama.

2. Pre-Processing Data

Pre-processing data merupakan tahap penting untuk memastikan kualitas dataset sebelum dilakukan pemodelan.

a. Pemilihan fitur atau parameter

Langkah awal dalam proses *pre-processing* adalah memilih kolom (fitur) yang relevan terhadap prediksi harga rumah. Ini membantu menghindari data yang tidak diperlukan dan memastikan model hanya fokus pada variabel yang memiliki pengaruh khususnya pada penelitian kali ini yang sudah ditentukan adalah luas bangunan (X1), luas tanah (X2), jumlah kamar tidur (X3), jumlah kamar mandi (X4), daya listrik (X5), kedekatan dengan jalan raya (X6) dan HARGA (Y).

b. Mengubah tipe data ke numerik.

Data yang diimpor dari Excel bisa saja terbaca dalam format string, khususnya jika terdapat simbol atau kesalahan entri. Oleh karena itu, setiap fitur numerik harus dikonversi menjadi tipe numerik (int atau float).

c. Menghapus data duplikat dan data yang mengandung nilai kosong.

Data duplikat dapat menyebabkan model menjadi bias karena memberikan bobot yang berlebihan terhadap data tertentu sehingga mengganggu akurasi prediksi. Sementara itu, keberadaan nilai kosong (NaN) dapat menimbulkan error saat proses pelatihan model maupun saat proses transformasi data.

d. Transformasi data kategorikal (*Label Encoding*)

Dalam penerapan model *Machine Learning* seperti *Linear Regression*, seluruh fitur masukan harus dalam bentuk numerik agar dapat dihitung secara matematis. Pada penelitian ini, kolom Dekat_jalan merupakan contoh fitur kategorikal biner yang dikonversi dengan memberi nilai 1 untuk "Ya" dan 0 untuk "Tidak". Variabel Dekat_jalan yang dimaksud merujuk pada kedekatan rumah dengan jalan raya utama sehingga rumah yang memiliki akses langsung atau berjarak dekat ke jalan raya utama diberi label "Ya" sedangkan yang lokasinya lebih jauh diberi label "Tidak".

e. Menghapus outlier (Z-Score)

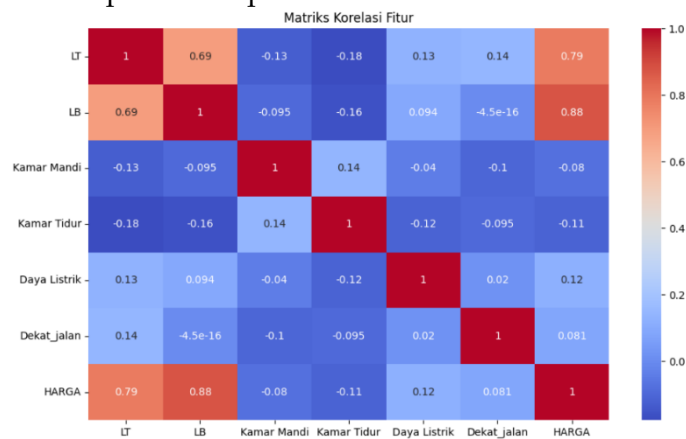
Teknik *Z-Score* pada tahap *pre-processing* dapat meningkatkan akurasi model dengan mentransformasikan data ke distribusi *zero-mean* dan *unit-variance* sehingga setiap fitur memiliki skala yang seragam dan lebih siap digunakan dalam regresi [9]. Rumus *Z-score* :

$$Z = \frac{X - \bar{X}}{\sigma}$$

di mana X = nilai data, \bar{X} = mean (rata-rata) data dan σ = standar deviasi.

Dalam penelitian ini, proses pendeteksian dilakukan dengan menghitung *Z-Score* untuk setiap fitur numerik, kemudian baris yang memiliki minimal satu nilai dengan *Z-Score* > 3 dihapus dari dataset.

Dari 7 variabel diatas saling terhubung atau memiliki korelasi. Korelasi dari setiap variabel atau fitur yang digunakan dapat dilihat pada Gambar 4.



Gambar 3. Metrik Korelasi

Berdasarkan hasil analisis korelasi pada Gambar 4 terlihat bahwa variabel luas bangunan (LB) memiliki korelasi tertinggi dengan harga rumah sebesar 0,88, diikuti oleh luas tanah (LT) dengan nilai korelasi 0,79. Hal ini menunjukkan bahwa kedua variabel tersebut merupakan faktor utama yang memengaruhi harga rumah di kawasan Kecamatan Paguyangan.

3. Pembagian Dataset

Pembagian data merupakan proses pembagian dataset menjadi dua bagian yaitu data latih (*training*) dan data uji (*testing*). Dalam penelitian ini digunakan teknik *K-Fold Cross Validation* sebanyak 10 *fold*. Metode *K-Fold Cross Validation* membagi dataset ke dalam k bagian yang sama (*fold*). Pada setiap iterasi, salah satu *fold* digunakan sebagai data uji (*testing*) sedangkan sisanya digunakan sebagai data latih (*training*).

Tabel 1 Pembagian Dataset

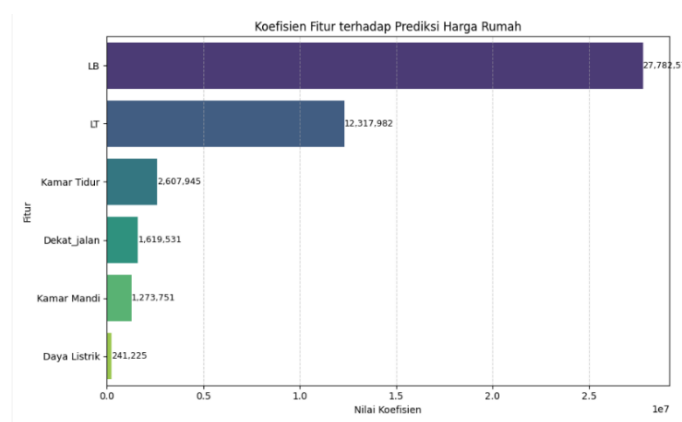
<i>Fold</i>	<i>Training</i>	<i>Testing</i>
1	2-10	1
2	1 dan 3-10	2
3	1-2 dan 4-10	3
4	1-3 dan 5-10	4
5	1-4 dan 6-10	5
6	1-5 dan 7-10	6
7	1-6 dan 8-10	7
8	1-7 dan 9-10	8
9	1-8 dan 10	9
10	1-9	10

4. Modelling Data

Pemodelan *Linear Regression* yang dilakukan dari 6 variabel antara lain luas bangunan, luas tanah, jumlah kamar tidur, jumlah kamar mandi, daya listrik dan kedekatan dengan jalan raya dengan masing-masing koefisien variabel dilihat pada Gambar 5 dan Tabel 2.

Tabel 2 Koefisien Fitur/Variabel

Variabel	Koefisien
X_1	27.782.579
X_2	12.317.982
X_3	1.273.751
X_4	2.607.945
X_5	241.225
X_6	1.619.531



Gambar 4. Nilai Koefisien Variabel

Berdasarkan nilai koefisien masing-masing variabel maka persamaan dari Linear Regression adalah sebagai berikut :

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \varepsilon$$

Dengan:

$X_1 = LB$, $X_2 = LT$, $X_3 = \text{Kamar Tidur}$, $X_4 = \text{Kamar Mandi}$, $X_5 = \text{Daya Listrik}$ dan $X_6 = \text{Dekat_jalan}$

Maka persamaan dalam penelitian ini adalah

$$Y = 261.301.412 + 27.782.579 \times LB + 12.317.982 \times LT + 2.607.945 \times \text{Kamar Tidur} + 1.273.751 \times \text{Kamar Mandi} + 241.225 \times \text{Daya Listrik} + 1.619.531 \times \text{Dekat_jalan}$$

5. Evaluasi Data

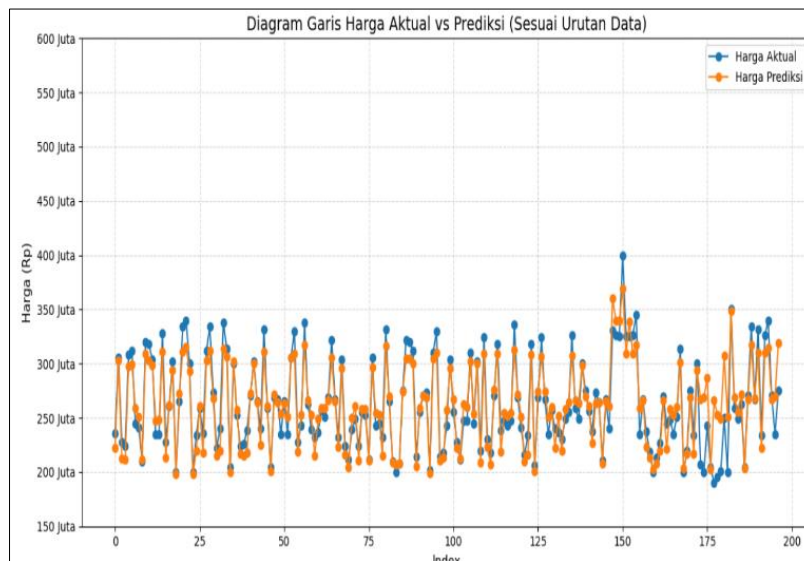
Evaluasi dilakukan secara kuantitatif terhadap hasil prediksi pada masing-masing *fold* dengan menggunakan beberapa metrik umum dalam regresi antara lain MSE (*Mean Squared Error*), RMSE (*Root Mean Squared Error*), MAE (*Mean Absolute Error*) dan R-squared (R^2). Hasil evaluasi tiap masing-masing *fold* adalah pada Gambar 6 sebagai berikut :

Fold	MSE	RMSE	MAE	R ²
1	Rp 261,294.326.884.832.09	Rp 16,164.601.04	Rp 11,330.471.56	0.8
2	Rp 421,085.216.647.473.38	Rp 20,520.361.03	Rp 13,015.605.30	0.7609
3	Rp 125,494.300.087.278.03	Rp 11,202.423.85	Rp 9,001.817.92	0.8819
4	Rp 181,303.161.405.242.06	Rp 13,464.886.24	Rp 10,271.285.78	0.862
5	Rp 147,632.962.551.964.50	Rp 12,150.430.55	Rp 10,589.444.81	0.8729
6	Rp 177,432.576.867.298.03	Rp 13,320.382.01	Rp 10,541.293.61	0.9427
7	Rp 611,310.483.235.149.25	Rp 24,724.693.79	Rp 17,561.243.53	0.7397
8	Rp 494,269.938.462.046.81	Rp 22,232.182.49	Rp 14,587.686.36	0.648
9	Rp 289,515.127.068.597.69	Rp 17,015.144.05	Rp 11,033.701.71	0.772
10	Rp 380,718.547.126.262.31	Rp 19,512.010.33	Rp 14,942.724.29	0.8027

Gambar 5. Evaluasi Data Setiap *Fold*

Hasil evaluasi dengan formula terbaik adalah fold ke-6 yang menunjukkan nilai *Mean Squared Error* (MSE) sebesar 177.432.576.867.298,03, *Root Mean Squared Error* (RMSE) sebesar 13.320.382,01, *Mean Absolute Error* (MAE) sebesar 10.541.293,61, dan koefisien determinasi (R²) sebesar 0,9427. Nilai RMSE yang berada pada kisaran Rp. 13.000.000,- dan MAE sebesar Rp. 10.000.000,- mengindikasikan bahwa prediksi harga rumah memiliki deviasi yang relatif kecil terhadap nilai aktual dengan melihat kisaran harga rumah di lokasi penelitian berada di ratusan juta. Selain itu, nilai R² sebesar 0,9427 menunjukkan bahwa sekitar 94,27% variasi harga rumah dapat dijelaskan oleh model berdasarkan variabel-variabel seperti luas tanah, luas bangunan, jumlah kamar tidur, jumlah kamar mandi, daya listrik, dan kedekatan dengan jalan utama.

Hasil ini menunjukkan bahwa model *Linear Regression* memiliki performa yang sangat baik dan mampu memberikan prediksi yang stabil untuk konteks skala lokal seperti di Kecamatan Paguyangan



Gambar 6. Grafik Harga Aktual dan Harga Prediksi

V. Kesimpulan dan Saran

1. Kesimpulan

Hasil evaluasi model yang dikembangkan dengan enam variabel utama antara lain luas bangunan, luas tanah, jumlah kamar tidur, jumlah kamar mandi, daya listrik, dan kedekatan dengan jalan raya menggunakan metode *10-Fold Cross Validation* menunjukkan bahwa formula terbaik diperoleh pada *fold* ke-6 dengan nilai koefisien determinasi (R²) sebesar 0,9427. Nilai ini menunjukkan bahwa sebesar 94,27% variasi harga rumah dapat dijelaskan oleh model regresi yang dibangun sedangkan sisanya dipengaruhi oleh faktor lain di luar model. Selain itu, model

juga dievaluasi menggunakan beberapa metrik regresi lainnya, yaitu *Root Mean Squared Error* (RMSE) sebesar Rp 13.320.382,01, *Mean Squared Error* (MSE) sebesar 177.432.576.867.298,03, dan *Mean Absolute Error* (MAE) sebesar Rp 10.541.293,61.

Nilai RMSE dan MAE yang relatif kecil menunjukkan bahwa rata-rata kesalahan prediksi harga rumah oleh model cukup rendah dan masih berada dalam batas wajar jika dibandingkan dengan kisaran harga rumah di kawasan Kecamatan Paguyangan yang mencapai ratusan juta rupiah.

Dengan demikian, dapat disimpulkan bahwa performa model *Linear Regression* dalam penelitian ini berada pada kategori sangat baik dan layak digunakan untuk konteks prediksi harga rumah. Keberhasilan model dalam menghasilkan prediksi yang akurat juga didukung oleh penerapan tahapan pra-pemrosesan data yang optimal, seperti transformasi fitur, penghapusan *outlier* menggunakan metode *Z-Score*, dan validasi model menggunakan teknik *10-Fold Cross Validation* untuk menghindari *overfitting*.

2. Saran

Beberapa saran yang dapat disampaikan untuk pengembangan penelitian dan implementasi di masa mendatang antara lain:

- 1) Menambahkan variabel-variabel independen lainnya yang lebih relevan untuk memperkaya model Regresi Linear dan menambahkan variasi data dari beberapa website jual-beli rumah [1].
- 2) Menerapkan algoritma prediksi lainnya yang lebih mendukung untuk melakukan prediksi secara lebih akurat [10].
- 3) Mengumpulkan lebih banyak data dari berbagai sumber untuk melatih model dengan informasi yang lebih komprehensif, termasuk data tambahan seperti kondisi lingkungan, aksesibilitas transportasi, dan fasilitas umum yang dapat meningkatkan akurasi model [11].

Referensi

- [1] A. Vermaysha and N. Nurmalitasari, "Prediksi Harga Rumah di Kabupaten Karanganyar Menggunakan Metode Regresi Linear," *Pros. Semin. Nas. Teknol. Inf. dan Bisnis*, pp. 6-11, 2023.
- [2] A. Widyastuti, "PREDIKSI HARGA RUMAH SESUAI SPESIFIKASI MENGGUNAKAN METODE MULTIPLE LINEAR REGRESSION," *SUBMIT*, vol. 4, no. 1, pp. 30-35, 2024.
- [3] M. Ridwan, T. H. Sinaga, and M. Elsera, "Penerapan Framework Codeigniter Dalam Perancangan Aplikasi Manajemen Iuran Perumahan Griya Mandiri," *Djtechno J. Teknol. Inf.*, vol. 3, no. 1, pp. 49-58, 2022, doi: 10.46576/djtechno.v3i1.2196.
- [4] A. A. Bastian, H. H. Handayani, and D. Wahiddin, "Implementasi Algoritma Support Vector Regression dan Linear Regression Untuk Prediksi Harga Rumah," *J. Ilm. Komput.*, vol. 20, pp. 961-972, 2024.
- [5] M. L. Mu'tashim, T. Muhayat, S. A. Damayanti, H. N. Zaki, and R. Wirawan, "Analisis Prediksi Harga Rumah Sesuai Spesifikasi Menggunakan Multiple Linear Regression," *Inform. J. Ilmu Komput.*, vol. 17, no. 3, p. 238, 2021, doi: 10.52958/iftk.v17i3.3635.
- [6] Nguyễn Quãn and an O. M. C. Safari, *Hands-On Application Development with PyCharm*, vol. 1. 2019.
- [7] S. Raschka, J. Patterson, and C. Nolet, "Machine learning in python: Main developments and technology trends in data science, machine learning, and artificial intelligence," *Inf.*, vol. 11, no. 4, 2020, doi: 10.3390/info11040193.

- [8] Romi Satria Wahono, "Research methodology," *Arch. Phys. Med. Rehabil.*, vol. 34, no. 4, pp. 217-224, 2017.
- [9] D. A. Anggoro, W. Supriyanti, and U. M. Surakarta, "Improving Accuracy by applying Z-Score Normalization in Linear Regression and Polynomial Regression Model for Real Estate Data," *Int. J. Emerg. Trends Eng. Res.* 7(11), Novemb. 2019, vol. 7, no. 11, 2019.
- [10] I. Maula, L. U. Hasanah, and A. Tholib, "Analisis Prediksi Harga Rumah Di Jabodetabek Menggunakan Multiple Linear Regression," *J. Inform. Kaputama*, vol. 7, no. 2, pp. 216-224, 2023, doi: 10.59697/jik.v7i2.135.
- [11] R. A. Saputra, A. Pratama, S. Informasi, U. Nahdlatul, and U. Kalimantan, "IMPLEMENTASI DECISION TREE UNTUK PREDIKSI HARGA RUMAH DI DAERAH TEBET Sistem Informasi STMIK Pontianak Abstraksi Pendahuluan Tinjauan Pustaka," vol. 6, no. 2, pp. 4-10, 2025.