

Penerapan Metode *Linear Regression* Untuk Memprediksi Harga Rumah

Muhammad Arizal Dwisakti¹, Muhammad Ridho Pramana², Riski Juliandri³ Taroni sokhi Zebua⁴,

¹Sistem Informasi, Fakultas : Ilmu Komputer, Universitas Budi darma, Medan, Indonesia

muhammadarizaldwisakti@gmail.com, mr8520219@gmail.com,

riskizuliandri966@gmail.com, taronizeb@gmail.com

Abstract - House prices are an important indicator in the property sector and are influenced by various factors, including physical characteristics of the building and location. This study aims to analyze and predict house prices using a linear regression method by utilizing several variables, namely land area, building area, number of bedrooms, number of bathrooms, parking availability, and distance to the city center. The data used in this study are secondary data collected through a web scraping process and are focused on houses with a price range of 300–700 million rupiah to represent the middle-market segment.

The research stages include data preprocessing, Pearson correlation analysis, multicollinearity testing, multiple linear regression modeling, and model performance evaluation using the coefficient of determination (R^2) and Root Mean Square Error (RMSE). The dataset is divided into 80% training data and 20% testing data. The results show that the constructed linear regression model achieves an R^2 value of 0.3078, indicating that the independent variables are able to explain 30.78% of the variation in house prices. The RMSE value of 117,482,242 indicates that prediction errors remain relatively high due to the wide variation in house prices.

The correlation analysis results reveal that the number of bathrooms and the distance to the city center have a relatively stronger relationship with house prices compared to other variables. This study demonstrates that linear regression can be used as an initial approach for house price prediction; however, it still has limitations in explaining overall price variations. Therefore, future research is expected to improve prediction performance by incorporating additional variables or applying more advanced modeling methods.

Keywords: house price, linear regression, coefficient of determination, RMSE, Pearson correlation.

Abstrak - Harga rumah merupakan salah satu indikator penting dalam sektor properti yang dipengaruhi oleh berbagai faktor, baik fisik bangunan maupun lokasi. Penelitian ini bertujuan untuk menganalisis dan memprediksi harga rumah menggunakan metode regresi linear dengan memanfaatkan variabel luas tanah, luas bangunan, jumlah kamar tidur, jumlah kamar mandi, tempat parkir, dan jarak ke pusat kota. Data yang digunakan dalam penelitian ini merupakan data sekunder yang dikumpulkan melalui proses web scraping dan difokuskan pada rumah dengan kisaran harga 100–700 juta rupiah untuk merepresentasikan segmen pasar menengah.

Tahapan penelitian meliputi prapemrosesan data, analisis korelasi Pearson, uji multikolinearitas, pemodelan regresi linear berganda, serta evaluasi kinerja model menggunakan koefisien determinasi (R^2) dan Root Mean Square Error (RMSE). Data dibagi menjadi data pelatihan sebesar 80% dan data pengujian sebesar 20%. Hasil penelitian menunjukkan bahwa

model regresi linear yang dibangun menghasilkan nilai R^2 sebesar 0,3078, yang menunjukkan bahwa variabel independen mampu menjelaskan sebesar 30,78% variasi harga rumah. Nilai RMSE sebesar 117.482.242 menunjukkan bahwa masih terdapat kesalahan prediksi yang dipengaruhi oleh variasi harga rumah yang cukup besar.

Hasil analisis korelasi menunjukkan bahwa jumlah kamar mandi dan jarak ke pusat kota memiliki hubungan yang relatif lebih kuat terhadap harga rumah dibandingkan variabel lainnya. Penelitian ini menunjukkan bahwa regresi linear dapat digunakan sebagai pendekatan awal dalam memprediksi harga rumah, namun masih memiliki keterbatasan dalam menjelaskan variasi harga secara menyeluruh sehingga diperlukan pengembangan model dengan penambahan variabel atau metode yang lebih kompleks pada penelitian selanjutnya.

Kata kunci: Harga rumah, Regresi linear, Koefisien Determinasi, RMSE, Korelasi Pearson.

1. PENDAHULUAN

Perkembangan sektor properti menjadi indikator penting bagi stabilitas ekonomi dan kesejahteraan masyarakat, terutama dalam hal akses terhadap tempat tinggal yang layak. Dalam beberapa penelitian, pendekatan data-driven menjadi sangat penting karena prediksi harga rumah mencerminkan dinamika pasar yang kompleks serta kebutuhan keputusan investasi yang tepat [1]. Kondisi ini memperlihatkan bahwa prediksi harga bukan sekadar pengukuran statistik tetapi juga bagian integral dari perencanaan sosial-ekonomi masyarakat modern.

Namun demikian, permasalahan utama yang dihadapi dalam prediksi nilai properti adalah variasi luas faktor yang saling memengaruhi, seperti lokasi, ukuran, jumlah kamar, dan kondisi bangunan yang berbeda-beda antara satu rumah dan rumahlain. Seperti yang dijelaskan dalam studi prediksi harga rumah dengan regresi linier yang dilakukan oleh Rosalia Roja Hallana & Fajri (2025), “prediksi harga rumah menjadi tantangan karena sifat data yang kompleks dan tergantung pada banyak variabel fisik serta kondisi pasar”[2] Hal ini menunjukkan bahwa metode statistik sederhana sering kali tidak cukup tanpa pendekatan pemodelan yang tepat dan pengelolaan data yang baik.

Sebagai sebuah solusi terhadap permasalahan variabilitas harga, metode *Linear Regression* banyak dipilih karena kemampuannya dalam menangkap hubungan linier antara variabel terkait dengan harga jual rumah. Misalnya, studi kasus prediksi harga rumah yang dikembangkan oleh Fathoni Dwi Atmoko (2023) menunjukkan bahwa model regresi linier mampu menjelaskan hingga 93,96 % variasi harga berdasarkan faktor-faktor seperti luas bangunan dan umur bangunan[3]. Pendekatan ini tidak hanya sederhana tetapi juga memberikan interpretasi yang mudah dipahami oleh para pemangku kepentingan seperti pembeli, penjual, dan analis pasar.

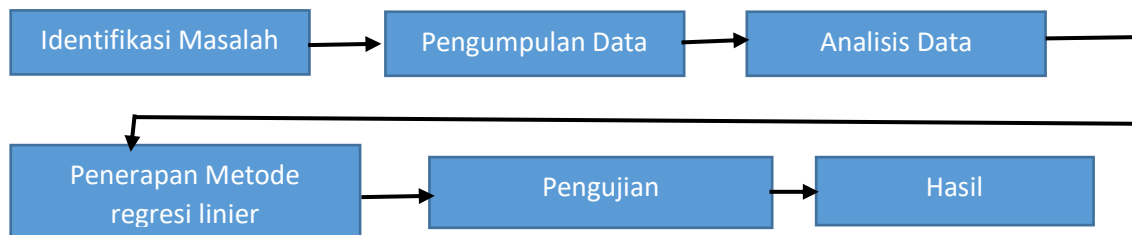
Manfaat dari penelitian ini bersifat praktis dan teoretis. Secara praktis, hasil model regresi linier dapat membantu pembeli dan penjual rumah dalam membuat keputusan berbasis estimasi harga yang objektif, mengurangi ketidakpastian dalam negosiasi dan perencanaan keuangan individu. Secara teoretis, penelitian ini memperkaya literatur prediksi harga rumah dengan memperlihatkan bagaimana model regresi linear bekerja secara empiris di data nyata, serta membuka peluang bagi pengembangan model lebih lanjut seperti regresi berganda atau algoritma *machine learning* lain yang lebih kompleks[4].

Tujuan dari penelitian ini adalah menerapkan metode *Linear Regression* untuk membangun model prediksi harga rumah, mengidentifikasi variabel paling berpengaruh terhadap harga, serta mengevaluasi kinerja model melalui metrik evaluasi seperti *R-squared* dan

RMSE. Dengan demikian, penelitian ini diharapkan dapat memberikan insight ilmiah yang kuat serta rekomendasi praktis bagi para pengambil keputusan di bidang properti dan teknologi data[2].

2. METODE

Untuk memprediksi harga rumah, pendekatan yang diambil dalam penelitian ini adalah melalui metode kuantitatif. Selanjutnya, pada fase metode penelitian yang akan dilakukan adalah pengumpulan informasi dengan variabel yang relevan, melakukan pemrosesan data awal (pre-processing data), kemudian menerapkan algoritma regresi linier pada dataset, dan hasil dari penelitian ini dapat dilihat pada gambar 1 dibawah ini.



Gambar 1. Tahap Metode Penelitian

3. HASIL DAN PEMBAHASAN

Dalam konteks prediksi harga rumah atau property, *Linear Regression* digunakan untuk memodelkan hubungan matematis antara satu atau lebih variabel independen dan variabel dependen sehingga memungkinkan prediksi variabel target berdasarkan informasi variabel lain[5]. Selain itu, perhitungan *Multiple Linear Regression* yang diterapkan dapat dinyatakan pada (1).

$$Y = a + b_1 X_1 + b_2 X_2 + \dots + b_n X_n \quad (1)$$

dengan

Y = Variabel dependen (variabel tak terikat)

$X_1, X_2, \dots, X_n$ = Variabel independent (variabel terikat)

a = Konstanta

$b_1, b_2, \dots, b_n$ = Koefisien regresi

Penelitian ini memanfaatkan platform Google Colab berbasis *web base* yang mempermudah proses melakukan analisis prediksi harga properti. Google Colab merupakan aplikasi web *open-source* dan juga memiliki berbagai fungsi lain dalam penelitian ini, seperti *data cleaning*, transformasi data, visualisasi data, *machine learning*, alat simulasi, dan berbagai kegunaan lainnya.

Tabel 1. Data Set sebanyak 500 data

No	Luas_Tanah	Luas_Bangunan	Kmr_Tidur	Kmr_Mandi	Tmpt_Parkir	Jrk_Ke_Kota	Harga
1	160	320	1	2	1	28	130000000
2	44	40	1	1	1	9	140000000
3	36	36	1	1	0	5	145000000
4	72	36	1	1	1	13	166000000

Penerapan Metode *Linear Regression* Untuk Memprediksi Harga Rumah

5	90	198	3	1	0	16	166000000
...	...	...	...	...	...	...	...
496	99	85	2	2	1	2	700000000
497	87	200	3	2	1	5	700000000
498	173	90	5	3	1	3	700000000
499	48	107	2	2	0	9	700000000
500	115	100	2	1	1	5	700000000

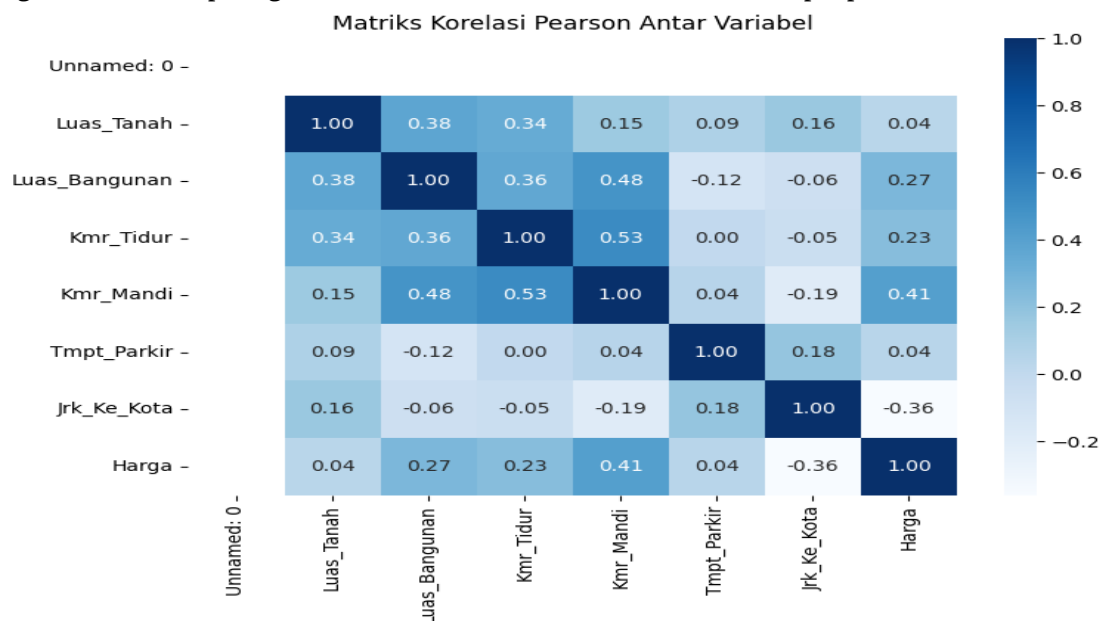
Hasil penerapan dengan metode *linear regression* menghasilkan hubungan antara luas tanah, luas bangunan, jumlah kamar tidur, jumlah kamar mandi, jumlah tempat parkir, jarak ke pusat kota dengan merujuk pada pernyataan (10).

(10)

$$Y = a + b_1(\text{luas tanah}) + b_2(\text{luas bangunan}) + b_3(\text{kamar tidur}) + b_4(\text{kamar mandi}) + b_5(\text{tempat parkir}) + b_6(\text{jarak ke kota})$$

Untuk mengetahui keterkaitan pada setiap variabel independen terhadap variabel dependen perlu adanya uji korelasi. Untuk menghitung nilai korelasi perlu mengetahui perhitungan nilai $r_{X_1Y}, r_{X_2Y}, r_{X_3Y}, r_{X_4Y}, r_{X_5Y}, r_{X_6Y}, r_{X_1X_2}, r_{X_1X_3}, r_{X_1X_4}, r_{X_1X_5}, r_{X_1X_6}$.

Pada gambar 3, terlihat tabel yang menunjukkan hubungan antara berbagai variabel. Dalam tabel ini, jelas bahwa faktor yang berpengaruh terhadap harga adalah luas bangunan, jumlah kamar mandi, hingga jarak ke kota. Sementara itu, korelasi yang tidak berpengaruh signifikan terhadap harga dalam set data tersebut adalah fasilitas tempat parkir mobil.



Gambar 3. Data korelasi dari setiap variabel

Koefisien determinasi dinyatakan dengan nilai R^2 yang memiliki rentang nilai antara 0 sampai 1. Semakin mendekati nilai 1, maka semakin besar kemampuan model dalam menjelaskan variasi data, sedangkan nilai yang mendekati 0 menunjukkan bahwa kemampuan penjelasan model relatif rendah.

Tabel 2. Hasil uji koefisien determinasi

Parameter	Nilai
R^2	0.3078

Berdasarkan hasil pengujian yang ditunjukkan pada Tabel 1, diperoleh nilai koefisien determinasi (R^2) sebesar 0,3078. Nilai tersebut menunjukkan bahwa variabel luas tanah, luas bangunan, jumlah kamar tidur, jumlah kamar mandi, tempat parkir, dan jarak ke pusat kota mampu menjelaskan sebesar 30,78% variasi harga rumah. Sementara itu, sebesar 69,22% variasi harga rumah dipengaruhi oleh faktor lain di luar variabel yang digunakan dalam penelitian ini.

Tabel 3. RMSE (Root mean square error)

Parameter	Nilai
RMSE	117482242.46

Selanjutnya, evaluasi kinerja model dilakukan menggunakan nilai Root Mean Square Error (RMSE). Nilai RMSE digunakan untuk mengukur rata-rata besarnya kesalahan prediksi antara nilai aktual dan nilai hasil prediksi. Semakin kecil nilai RMSE, maka semakin baik kinerja model dalam melakukan prediksi harga rumah. Nilai RMSE yang diperoleh sebesar 117482242.46 atau Rp. 117.482.242,46.

4. PENUTUP

4.1. Kesimpulan

Penelitian ini menunjukkan bahwa metode linear regression dapat digunakan untuk memprediksi harga rumah pada rentang harga 100–700 juta rupiah dengan memanfaatkan variabel luas tanah, luas bangunan, jumlah kamar tidur, jumlah kamar mandi, tempat parkir, dan jarak ke pusat kota. Hasil evaluasi model menghasilkan nilai koefisien determinasi (R^2) sebesar 0,3078 yang menunjukkan bahwa variabel independen mampu menjelaskan sekitar 30,78% variasi harga rumah, sementara nilai RMSE yang relatif tinggi disebabkan oleh variasi harga rumah yang cukup besar. Analisis korelasi menunjukkan bahwa jumlah kamar mandi dan jarak ke pusat kota memiliki hubungan yang lebih kuat terhadap harga rumah dibandingkan variabel lainnya. Secara keseluruhan, regresi linear dapat digunakan sebagai pendekatan awal dalam prediksi harga rumah, namun diperlukan penambahan variabel dan pengembangan metode untuk meningkatkan kinerja model pada penelitian selanjutnya

4.2. Saran

Berdasarkan hasil penelitian yang telah dilakukan, disarankan agar penelitian selanjutnya menambahkan variabel lain yang lebih relevan, seperti kondisi lingkungan, akses fasilitas umum, dan kualitas bangunan, untuk meningkatkan kemampuan model dalam menjelaskan variasi harga rumah. Selain itu, penggunaan metode pemodelan yang lebih kompleks, seperti machine learning, dapat

dipertimbangkan untuk memperoleh hasil prediksi yang lebih akurat. Penelitian selanjutnya juga disarankan untuk menggunakan jumlah data yang lebih besar dan rentang wilayah yang lebih luas agar hasil yang diperoleh lebih representative.

DAFTAR PUSTAKA

- [1] H. Hakim, D. Kamil, and B. Alatas, "Pendekatan Machine Learning untuk Estimasi Harga Rumah dengan Regresi Linier," *ALPHA J. Sci. Technol.*, vol. 1, no. 1, pp. 18–22, Jan. 2025, doi: 10.70716/alpha.v1i1.99.
- [2] R. R. Hallan and I. N. Fajri, "Prediksi Harga Rumah menggunakan Machine Learning Algoritma Regresi Linier," *J. Teknol. Dan Sist. Inf. Bisnis*, vol. 7, no. 1, pp. 57–62, 2025, doi: 10.47233/jteksis.v7i1.1732.
- [3] Fathoni Dwi Atmoko, "Prediksi Harga Rumah dengan Regresi Linier," *Polyg. J. Ilmu Komput. dan Ilmu Pengetah. Alam*, vol. 1, no. 2, pp. 01–10, Mar. 2023, doi: 10.62383/polygon.v1i2.854.
- [4] G. A. Syafarina, "Implementasi Framework Streamlit Sebagai Prediksi Harga Jual Rumah Dengan Linear Regresi lain menjelaskan bahwa Pegujian akurasi dilakukan dengan mengukur Root Mean Squared Error (RMSE). Nilai rata-rata jual rumah yang akurat di Kota Banjarmasin , mengg," pp. 121–125, 2023, doi: 10.47002/metik.v7i2.680.
- [5] S. Tania and Albertus Dwiyoga Widianoro, "Implementasi Metode Regresi Linier untuk Analisis Prediksi Harga Saham TSLA di RapidMiner," *Sist. J. Ilm. Sist. Inf.*, vol. 1, no. 1, pp. 27–35, Jul. 2024, doi: 10.69533/67a7h013.