



Analisis Prediksi Harga Rumah di Bandung Menggunakan Regresi Linear Berganda

Rafif Nauval Tuah Siregar

Universitas Negeri Medan

Korespondensi penulis: rafifnaufal37@email.com

Vijay Sitorus

Universitas Negeri Medan

Willy Pramudia Ananta

Universitas Negeri Medan

Alamat: Jl. William Iskandar Ps. V, Kenangan Baru, Kec. Percut Sei Tuan, Kabupaten Deli Serdang, Sumatera Utara 20221

Abstract. This research aims to develop a model for estimating house prices in the Bandung area using the Multiple Linear Regression approach. House prices play a significant role in the decision-making process for purchases. The ability of this model to predict house prices with high accuracy provides significant benefits for potential buyers, sellers, and various stakeholders in the housing industry. Data on house prices and potential variables such as land area, building area, number of bedrooms, nearby facilities, and geographical location were collected for analysis. The use of Multiple Linear Regression allows for a deeper understanding of the relationships between these variables and the value of the house. The analysis results show a strong correlation between these variables and house prices in Bandung. The developed Multiple Linear Regression model can provide satisfactory predictions of house prices. This model can be used as a tool for both homebuyers and sellers to determine fair prices and assist property developers in identifying key factors influencing house prices in the Bandung region.

Keywords: Prediction, House Price, Multiple Linear Regression

Abstrak. Penelitian ini memiliki tujuan untuk mengembangkan model memperkirakan harga rumah di daerah Bandung dengan menerapkan pendekatan Regresi Linear Berganda. Harga rumah berperan sangat signifikan dalam tahap pengambilan keputusan untuk melakukan pembelian. kemampuan model ini dalam memprediksi harga rumah dengan akurasi tinggi memberikan manfaat besar bagi calon pembeli, penjual, dan berbagai pihak yang terlibat dalam industri perumahan. Data harga rumah dan variabel-variabel potensial seperti luas tanah, luas bangunan, dan jumlah kamar tidur, fasilitas terdekat, dan lokasi geografis dikumpulkan untuk dianalisis. Penggunaan metode Regresi Linear Berganda memungkinkan pemahaman yang lebih mendalam mengenai keterkaitan antara variabel-variabel tersebut dan nilai rumah. Hasil analisis menunjukkan terdapat hubungan yang cukup kuat antara variabel-variabel tersebut dengan harga rumah di Bandung. Model Regresi Linear Berganda yang dikembangkan mampu memberikan prediksi harga rumah dengan tingkat akurasi yang memuaskan. Model ini dapat digunakan sebagai alat bantu bagi pembeli dan penjual rumah dalam menentukan harga yang adil, sekaligus membantu pengembang properti dalam mengidentifikasi faktor-faktor utama yang memengaruhi harga rumah di wilayah Bandung.

Kata kunci: Prediksi, Harga rumah, Regresi Linear Berganda

LATAR BELAKANG

Rumah menjadi kebutuhan pokok yang tidak dapat dihindari oleh masyarakat, berfungsi sebagai tempat perlindungan dan istirahat dari kesibukan sehari-hari. Mirip dengan emas, rumah juga dapat dianggap sebagai bentuk investasi untuk masa depan. Hal ini dikarenakan adanya fluktuasi harga yang dapat berubah sewaktu-waktu dan permintaan yang terus meningkat, terutama jika properti tersebut berlokasi secara strategis, dekat dengan area kerja, pusat perkantoran, pusat perbelanjaan, fasilitas transportasi, dan sebagainya. Semua faktor ini tentu akan memberikan dampak yang signifikan terhadap kenaikan harga rumah (Andi, et al., 2021).

Sejalan dengan berlalunya waktu, kebutuhan fisiologis manusia semakin meningkat, termasuk keinginan untuk memiliki rumah. Pengembang properti bersaing untuk membangun properti, khususnya rumah, sebagai bentuk investasi. Fenomena ini menyebabkan harga rumah naik seiring dengan peningkatan daya beli masyarakat. Dalam situasi ini, calon pembeli rumah perlu menilai apakah properti yang akan dibeli akan memberikan keuntungan atau tidak. Dalam konteks investasi, analisis yang teliti sangat diperlukan agar dapat memprediksi pergerakan harga dengan akurat dan menghindari potensi kerugian (Halimi, et al., 2019).

Dalam studi dengan judul "Membangun model penilaian real estate dengan pendekatan perbandingan melalui penalaran berbasis kasus," yang dilakukan oleh I-Cheng Yeh, Tzu-Kuang Hsu (Yeh, dkk., 2018), tujuannya adalah mengajukan metode penilaian real estate yang inovatif dengan kemampuan menghitung koefisien koreksi. Penelitian ini menggunakan lima model regresi, yakni Model Linier, Model Kuadrat, Model Logaritmik, Model Eksponensial, dan Model Pertumbuhan Eksponensial. Hasil dari pendekatan perbandingan kuantitatif menunjukkan tingkat kesalahan hit rate sebesar 20%, dengan skor RMSE sebesar 8,65 (Rahayuningtyas, et al., 2021).

Dengan demikian, dalam penelitian ini, peneliti memutuskan untuk memanfaatkan pendekatan Regresi Linear Berganda dengan memperhitungkan faktor-faktor yang berkaitan dengan faktor-faktor yang mempengaruhi prediksi harga rumah, dengan harapan mendapatkan prediksi yang lebih akurat. Metode Regresi Linear Berganda diharapkan dapat menghasilkan output yang sesuai dengan pola data training.

KAJIAN TEORITIS

1. Rumah

Secara keseluruhan, suatu tempat tinggal dapat digambarkan sebagai area yang memberikan perlindungan dan tempat berlindung bagi sebuah keluarga dari beragam kondisi di lingkungan sekitarnya, termasuk paparan sinar matahari yang intens, cuaca buruk seperti badai atau hujan, hingga suhu dingin pada malam hari. Lebih dari itu, tempat tinggal juga memiliki peran sebagai representasi dari identitas individu dalam usaha meningkatkan kualitas hidup, serta membentuk sifat, karakter, dan kepribadian suatu bangsa (Sanusi, et al.,2020).

Dalam konteks fisiknya, rumah dapat diidentifikasi sebagai struktur atau lokasi di mana keluarga dapat berkumpul, bekerja, serta berfungsi sebagai area untuk beristirahat dan mendapatkan kenyamanan fisik dan mental setelah menjalankan berbagai aktivitas harian masing-masing bagi para penghuni. Pandangan ini diperkuat oleh perspektif Tuner, yang mengungkapkan bahwa rumah memiliki peran utama sebagai pendukung rasa aman, menciptakan peluang, dan membentuk identitas keluarga. Dari segi psikologis, rumah menjadi lingkungan yang menciptakan suasana yang nyaman, aman, dan penuh kegembiraan bagi semua penghuninya (Adhyakas,2019).

2. Faktor yang mempengaruhi nilai jual rumah

Selain mempertimbangkan nilai tanah, terdapat sejumlah faktor lain yang turut memengaruhi penetapan harga jual rumah. Disarankan bagi calon pembeli untuk melakukan evaluasi terhadap atribut-atribut bangunan, melibatkan luas bangunan, jumlah kamar tidur, jumlah kamar mandi, dan kondisi keseluruhan rumah yang akan dibeli. Selanjutnya, disarankan pula untuk melakukan inspeksi menyeluruh terhadap keunggulan dan kelemahan properti rumah, seperti kondisi plafon dan rangka bangunan. Selain itu, perlu untuk memeriksa apakah rumah yang akan dibeli mengalami kerusakan dan memerlukan perbaikan. Setelah itu, melakukan pembandingan harga rumah dengan harga pasar dan mempertimbangkan lokasi hunian, terutama jika rumah tersebut berada di lokasi strategis, dekat dengan berbagai fasilitas, dan memiliki akses yang mudah. Semakin baik kualitas rumah dan semakin strategis lokasinya, maka harga rumah kemungkinan akan lebih tinggi (Kurniawan,2019).

3. Regresi linear berganda

Model Regresi Linear Berganda melibatkan lebih dari satu variabel, di mana model ini Melakukan prediksi dengan menggunakan data yang memiliki skala interval atau rasio, dan mampu menghasilkan lebih dari satu prediksi, merupakan esensi dari metode ini.

Pendekatan ini melibatkan analisis variabel independen dan variabel dependen. (Skillplus, 2019). Persamaan untuk model regresi linear berganda dapat diungkapkan sebagai berikut:

$$Y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \cdots + \beta_nx_n + e$$

Ket:

Y = Variabel Tergantung (Dependen)

x_1, x_2, \dots, x_n = Variabel independen

a = Konstanta (Intercept)

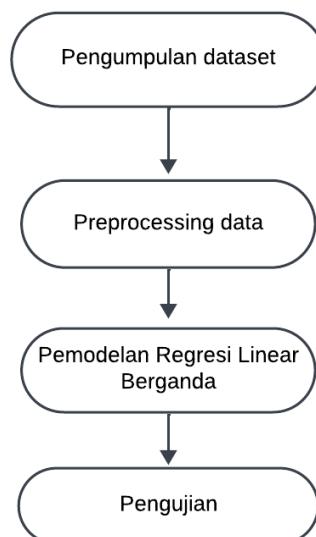
β_0 = intercept (konstanta), yaitu nilai y ketika semua x bernilai 0.

$\beta_1, \beta_2, \dots, \beta_n$ = koefisien regresi yang mengukur seberapa besar perubahan dalam y sebagai akibat sebagai akibat dari perubahan satu unit dalam variabel independen (x_1, x_2, \dots, x_n).

e = Error

METODE PENELITIAN

Pada tahap ini, terdapat serangkaian langkah yang dilakukan oleh penulis dalam pelaksanaan penelitian ini. Langkah-langkah tersebut mencakup pengumpulan dataset, kemudian masuk ke tahap preprocessing data, dilanjutkan dengan langkah-langkah untuk menghilangkan noise data, dan selanjutnya melibatkan melakukan pemodelan dengan memanfaatkan General Regression Neural Network. Tahap terakhir melibatkan sejumlah uji akurasi sebagai penilaian terhadap kinerja model.



Gambar 1. Alur penelitian

A. Pengumpulan Dataset

Dalam rangka penelitian ini, peneliti memanfaatkan dataset yang diperoleh dari Kaggle, yakni dataset mengenai harga rumah di Bandung (<https://www.kaggle.com/datasets/raflaping/dataset-harga-rumah-bandung>). Dataset ini terdiri dari kelas-kelas yang terkait dengan atribut-atribut yang digunakan untuk memprediksi harga rumah, antara lain jumlah kamar (kamar), luas tanah (lahan), luas bangunan (bangunan), serta beberapa atribut tambahan yang tidak terlalu krusial seperti judul, deskripsi, dan alamat. Jumlah total data dalam dataset ini mencapai 1470.

B. Preprocessing Data

Proses preprocessing memegang peranan krusial dengan melibatkan serangkaian langkah-langkah sebelum data diinput ke dalam model atau algoritma. Fungsinya adalah untuk menyiapkan dan meningkatkan kualitas atau kesiapan data agar sesuai dengan kebutuhan dalam melakukan prediksi harga rumah.

C. Pemodelan Regresi linear berganda

Pada tahap ini, peneliti melakukan pemodelan dataset harga rumah Bandung menggunakan Regresi Linear Berganda. Variabel dependen dalam model ini adalah harga, sedangkan variabel independennya mencakup jumlah kamar, luas lahan, dan luas bangunan. Dalam konteks regresi linear berganda, hubungan ini dijelaskan sebagai garis linear. Model regresi linear berganda bisa diungkapkan dalam bentuk persamaan matematika berikut:

$$Y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \cdots + \beta_nx_n + e$$

D. Pengujian

Berikutnya, dalam tahap terakhir, peneliti melakukan pengujian untuk mengevaluasi tingkat akurasi prediksi dan kinerja dari model Regresi Linear Berganda.

HASIL DAN PEMBAHASAN

A. Data Preprocessing

Dataset ini memiliki total 1470 baris dan 7 kolom. Langkah awal melibatkan pembersihan data dari nilai yang hilang (missing value) dan atribut yang tidak relevan, seperti atribut nomor, judul, alamat, dan deskripsi. Hasil penelitian menunjukkan bahwa terdapat 8 baris dengan nilai 0 pada kolom kamar, dan masing-masing terdapat 2 baris dengan nilai 0 pada kolom bangunan dan lahan. Baris-baris ini dieliminasi karena tidak mungkin memiliki nilai 0. Sehingga, dataset telah diperbaiki dan siap untuk tahap pemrosesan selanjutnya.

Table 1. hasil dari cleaning data dan penghapusan atribut yang tidak dipakai

Kamar	bangunan	lahan	harga
3	75	84	590000000
3	70	71	700000000
2	36	72	386000000
2	60	60	499000000
2	90	114	397500000

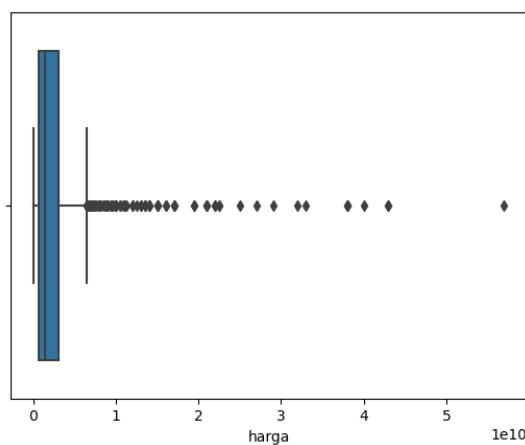
Data di atas adalah sebagian teratas dari dataset yang telah melewati proses pembersihan, di mana atribut-atribut yang tidak diperlukan, seperti nomor, judul, alamat, dan deskripsi, telah dihilangkan. Data ini sekarang siap untuk diproses pada langkah berikutnya.

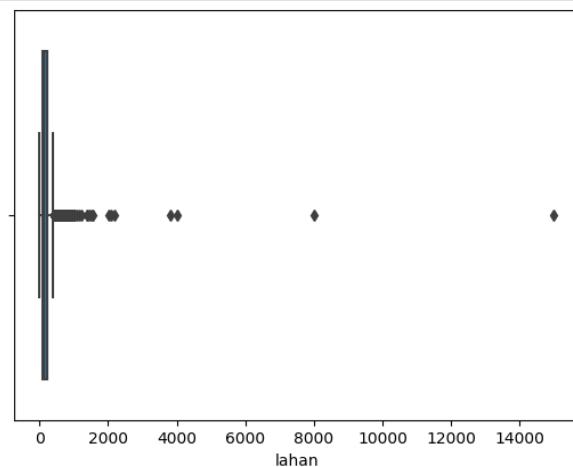
Penjelasan setiap kolom:

- Kamar : jumlah kamar pada rumah
- Bangunan: luas bangungan rumah
- Lahan : luas lahan
- Harga : harga rumah tersebut

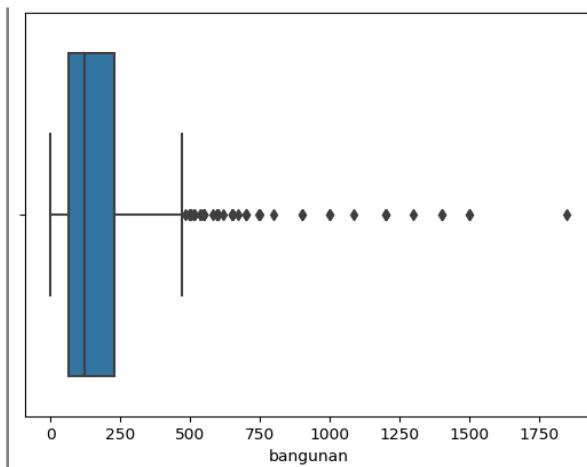
B. Eksplorasi Data

Tahap selanjutnya melibatkan penemuan dan penghapusan outlier pada setiap atribut data rumah. Berikut adalah representasi visual melalui boxplot untuk masing-masing atribut yang menunjukkan adanya outlier.

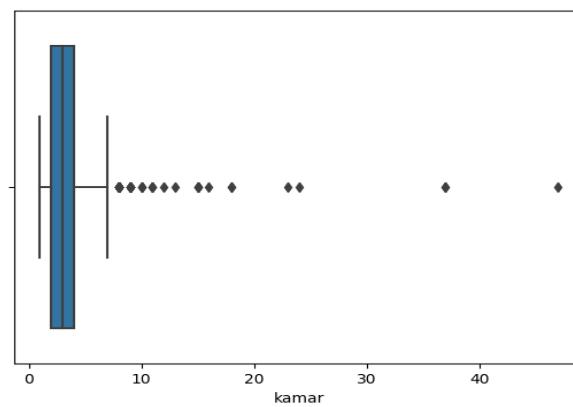
**Gambar 2.** boxplot atribut harga



Gambar 3. boxplot atribut lahan



Gambar 4. boxplot atribut bangunan



Gambar 5. boxplot atribut kamar

C. Splitting Data

Data tersebut telah melalui tahap pembersihan, dan selanjutnya, dilakukan pembagian dataset menjadi dua bagian, yaitu data latih dan data uji. Dalam tahap ini, data dibagi dengan proporsi 80% untuk data latih dan 20% untuk data uji. Penentuan

proporsi ini didasarkan pada jumlah data yang mencukupi, sehingga pemilihan proporsi 80:20 dianggap lebih efektif dibandingkan dengan 70:30.

D. Hasil

Proses pengembangan model Metode Regresi Linear Berganda diawali dengan membagi dataset menjadi dua kelompok, yakni set data pelatihan dan set data pengujian. Tiga variabel dependen yang digunakan adalah jumlah kamar (kamar), luas bangunan (bangunan), dan luas lahan (lahan), sedangkan variabel independennya adalah harga setiap rumah. Data latih, yang mencakup 80% dari total data, digunakan untuk Proses pelatihan model Regresi Linear Berganda dilakukan pada saat yang bersamaan, sementara data pengujian dimanfaatkan untuk mengukur performa algoritma ini. dapat memberikan hasil yang akurat. Evaluasi tersebut mencakup pengukuran Mean Absolute Error (MAE), Mean Squared Error (MSE), dan akurasi model, yang hasilnya dapat dilihat sebagai berikut :

Table 2. Hasil Pengujian

Mean Absolute Error (MAE)	458447967.3541158
Mean Squared Error (MSE)	4.20480881505
Model Accuracy	0.7729522512

Dari evaluasi yang dilakukan, dapat dilihat bahwa tingkat akurasi yang berhasil dicapai sebesar 0.77 atau 77%, sebuah hasil yang cukup memuaskan untuk memprediksi Menentukan nilai harga suatu rumah berdasarkan variabel independen yang ada. Selanjutnya, tingkat akurasi dari hasil ini akan diuji melalui proses prediksi harga untuk sebuah rumah dengan spesifikasi tertentu.

- Jumlah kamar : 2
- Luas bangunan : 70
- Luas tanah : 90

```
fitur_baru = [[2, 70, 90]] |  
harga_prediksi = regressor.predict(fitur_baru)  
print("Prediksi Harga Rumah:", harga_prediksi[0])  
  
Prediksi Harga Rumah: 799002724.7196944
```

Gambar 6. prediksi harga rumah sesuai dengan spesifikasi

Perkiraan harga berdasarkan spesifikasi di atas adalah Rp 799.002.724 (799 Juta Rupiah).

KESIMPULAN DAN SARAN

Prediksi harga rumah dapat Terpengaruh oleh sejumlah faktor, seperti dimensi lahan, luas bangunan, dan jumlah kamar tidur. Dalam usaha untuk memprediksi harga berdasarkan variabel-variabel ini, diperlukan metode yang dapat mendukung, dan salah satu metode yang relevan adalah Multiple Linear Regression. Prediksi harga dapat dihasilkan dengan mempertimbangkan berbagai faktor pada variabel yang ada. Penting untuk memilih faktor-faktor tersebut sesuai dengan kriteria tertentu agar prediksi yang dihasilkan memiliki tingkat akurasi yang tinggi. Dalam pengujian menggunakan sampel data sebanyak 1470 baris Dengan 7 kolom yang memuat informasi tentang harga rumah di Bandung, setelah melalui tahap pembersihan data, langkah selanjutnya adalah memisahkan antara dataset pelatihan dan pengujian. Dari dataset tersebut, dilakukan perhitungan akurasi menggunakan Model Regresi Linear Berganda, mencapai tingkat keakuratan sebesar 77%. Hasil ini dianggap memadai untuk melakukan prediksi harga rumah sesuai dengan spesifikasi yang diinginkan.

DAFTAR REFERENSI

- Adhyaksa. (2019, May 6). Definisi Rumah Sebagai sebuah Bangunan untuk Tempat Tinggal. Diambil kembali dari <https://www.adhyaksapersada.co.id/pengertian-rumah/>
- Halimi, I., Azhar, Y., & Marthasari, G. I. (2019). Prediksi Harga Emas Menggunakan Univariate Convolutional Neural Network. *Jurnal Reppositor*, 105-116.
- Kurniawan, Y. D. (2021). ANALISIS FAKTOR-FAKTOR YANG MEMPENGARUHI PEMBELIAN RUMAH DI KOTA MADIUN. *J-MACC, Journal of Management and Accounting*, 72-85.
- Rahayuningtyas, E. F., Rahayu, F. N., & Azhar, Y. (2021). Prediksi Harga Rumah Menggunakan General Regression Neural Network. *Jurnal Informatika*, 59-66.
- Saiful, A., Andryana, S., & Gunarryati, A. (2021). Prediksi Harga Rumah Menggunakan Web Scrapping Dan Machine Learning Dengan Algoritma Linear Regression. *Jurnal Teknik Informatika dan Sistem Informasi*, 41-50.
- Sanusi, R. M., Ansori, A. S., & Wijaya, R. (2020). Prediksi Harga Rumah Di Kota Bandung Bagian Timur Dengan Menggunakan Metode Regresi. *e-Proceedings of Engineering*, 9381-9395.
- Skillplus. (2019, Agustus 2). Multiple Linear Regression-pendahuluan. Diambil kembali dari <https://skillplus.web.id/multiple-linear-regression-pendahuluan/>
- Yeh, I.-C., & Hsu, T.-K. (2018). Building real estate valuation models with comparative approach through case-based reasoning. *Applied Soft Computing*, 260-271.