

Analisis Prediksi Harga Rumah Pada Machine Learning Menggunakan Metode Regresi Linear

Nuzuliarini Nuris
Universitas Bina Sarana Informatika
nuzuliarini.nzn@bsi.ac.id

Abstrak - Rumah merupakan kebutuhan esensial yang sangat penting bagi setiap individu, dengan pentingnya peran rumah tersebut maka seseorang yang ingin membeli rumah dan seseorang yang ingin menjual rumah sangat membutuhkan prediksi harga rumah yang sesuai dan akurat. Dalam menentukan prediksi harga rumah ditentukan oleh banyak faktor. Berdasarkan permasalahan tersebut maka setelah dilakukan penelitian dengan menggunakan metode regresi linear untuk memprediksi harga rumah sehingga dapat mempermudah transaksi jual dan beli. Data yang digunakan dalam penelitian ini adalah data sekunder yang bersumber dari website rumah123.com. Hasil penelitian ini menunjukkan nilai akurasi sebesar 86,41%. Harga rumah sangat berpengaruh kuat dengan luas bangunan dan harga rumah tidak dipengaruhi pada kriteria jenis sertifikat dikarenakan hasil yang didapat pada keterkaitan atau hubungan variabel luas bangunan sebesar 0,80 dan jenis sertifikat sebesar 0,00. Model dievaluasi menggunakan metrik evaluasi yang sesuai dengan menguji akurasi dan kinerja dari mean squared error (MSE) sebesar 464,87397, R-squared (R^2) sebesar 0,7818811 dan mean absolute error (MAE) sebesar 313,00083.

Kata kunci: regresi linear, prediksi harga rumah

Abstract - A house is an essential need that is very important for every individual, with the important role of the house, someone who wants to buy a house and someone who wants to sell a house really needs appropriate and accurate house price predictions. In determining house price predictions, many factors are determined. Based on these problems, research was carried out using the linear regression method to predict house prices so that buying and selling transactions could be made easier. The data used in this research is secondary data sourced from the rumah123.com website. The results of this research show an accuracy value of 86.41%. House prices have a very strong influence on the building area and house prices are not influenced by the certificate type criteria because the results obtained from the relationship between the building area variables are 0.80 and the certificate type is 0.00. The model was evaluated using appropriate evaluation metrics by testing the accuracy and performance of the mean squared error (MSE) of 464.87397, R-squared (R^2) of 0.7818811 and mean absolute error (MAE) of 313.00083.

Keyword : linear regression, house price prediction

1. Pendahuluan

Rumah merupakan kebutuhan esensial yang sangat penting bagi setiap individu [1]. Dengan pentingnya peran rumah dalam kehidupan masyarakat, banyak orang yang berkeinginan untuk membeli atau menjual rumah [2]. Saat seseorang memutuskan untuk memberi dan berinvestasi dalam properti pada sebuah bangunan rumah, sangat perlu mengetahui harga rumah berdasarkan dengan kriteria yang diinginkan. Selain sandang dan pangan, rumah merupakan kebutuhan dasar manusia [3]. Seiring waktu, rumah juga berkembang menjadi salah satu bentuk investasi yang menarik [4]. Hal ini menjadikan daftar harga rumah yang beragam sesuai dengan faktor utama yang biasanya dipertimbangkan dalam harga rumah, sehingga dapat memungkinkan

untuk dapat memprediksi rumah berdasarkan faktor yang disediakan [5]. Properti adalah salah satu bentuk investasi yang memiliki potensi return yang tinggi [6]. Tujuan dari prediksi harga rumah tidak hanya bermanfaat bagi para pembeli dan penjual, tetapi juga bagi pengembang, agen properti, dan investor [7]. Para pengusaha properti akan bersaing untuk membangun atau mengakuisisi rumah sebagai investasi [8]. Menurut penjual dapat menggunakan prediksi harga untuk menetapkan harga jual yang kompetitif dan menguntungkan. Oleh karena itu dibutuhkan sebuah prediksi yang dapat menghasilkan akurasi data yang tinggi sehingga didapatkan menyesuaikan jumlah produksi sesuai dengan kebutuhan konsumen [9].

Memprediksi harga rumah menjadi sangat penting [10] karena dapat memberikan panduan yang berguna bagi calon pembeli, penjual, dan pihak terkait lainnya dalam mengambil keputusan yang lebih baik dan berdasarkan informasi yang tepat [11]. Ketidakpastian dan ketidakdapatannya memprediksi harga membuat investor atau pembeli rumah mencari sistem yang dapat meramalkan harga rumah [12]. Dalam machine learning, studi ini termasuk dalam kategori supervised learning, di mana teknik pemodelan ini menggunakan data yang telah dilabeli atau memiliki target tertentu. Tujuannya adalah untuk mengidentifikasi hubungan kausal antara variabel input (variabel independen) dan variabel targetnya (variabel y) [13]. Regresi linear adalah suatu teknik statistik yang digunakan untuk menemukan dan menggambarkan hubungan antara variabel yang bergantung (seperti harga rumah) dengan satu atau lebih variabel independen [14]. Variabel-variabel seperti luas tanah, luas bangunan, jumlah kamar tidur, jumlah kamar mandi, dan daya listrik memiliki pengaruh yang signifikan terhadap harga rumah di wilayah Bekasi. Dengan menerapkan regresi linear [15], kita dapat mengeksplorasi dan menemukan pola serta hubungan yang ada di antara variabel-variabel tersebut dengan harga rumah di wilayah Bekasi.

2. Kajian Pustaka

a. Penelitian Terdahulu

- 1) Melakukan prediksi harga rumah dengan metode Neural Network dengan variabel seperti luas tanah, jumlah kamar tidur, dan jumlah kamar mandi memengaruhi harga rumah. akurasi dan evaluasi kinerja model dengan hasil mencapai tingkat keakuratan 74% [16].
- 2) Melakukan Klasifikasi Menggunakan Algoritma Support Vector Machine (SVM) yang diterapkan dengan menggunakan berbagai kernel, dan kernel Radial Basis Function (RBF) menunjukkan performa yang optimal. Kernel ini berhasil menjelaskan sebanyak 70,78% variasi harga rumah [5].
- 3) Dengan menggunakan Metode random forest berhasil memprediksi harga rumah dengan tingkat akurasi mencapai 75,10% [7].

b. Machine learning

Machine learning adalah bagian dari kecerdasan buatan yang fokus pada pengembangan teknik di mana komputer dapat belajar dari data dan pengalaman, tanpa

perlu diprogram secara eksplisit. *machine learning* memungkinkan komputer untuk mengidentifikasi pola dalam data, melakukan prediksi, dan mengambil keputusan dengan sedikit atau tanpa campur tangan manusia.

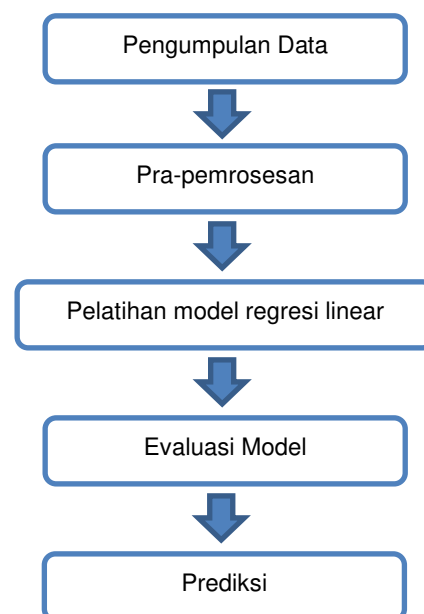
c. Algoritma Regresi Linear

Regresi Linear adalah metode dalam statistik dan *machine learning* yang digunakan untuk memahami hubungan antara variabel dependen (target) dengan satu atau lebih variabel independen (*predictor*). Tujuannya adalah untuk menemukan garis atau pola linear yang paling cocok untuk menjelaskan hubungan antara variabel-variabel tersebut.

Pada dasarnya, Regresi Linear berusaha mencari garis lurus atau *hyperplane* yang paling sesuai dengan data yang diamati. Garis atau *hyperplane* ini kemudian digunakan untuk membuat prediksi berdasarkan nilai-nilai variabel independen yang diberikan.

3. Metode Penelitian

Dalam melakukan prediksi harga rumah, pendekatan penelitian melibatkan metode kuantitatif. Langkah-langkah penelitian ini mencakup pengumpulan data yang melibatkan variabel terkait, pemrosesan dan pembersihan data, melakukan eksplorasi data, menerapkan algoritma regresi linear pada data, dan akhirnya menganalisis hasil penelitian. Berikut gambaran langkah penelitian;



Gambar 1. Langkah-langkah Penelitian

1. Pengumpulan Data

Pada tahap ini data yang digunakan ialah data yang tersedia pada laman website rumah123.com. Isi data berupa data harga rumah yang berada pada wilayah daerah

bekasi yang sedang dipasarkan oleh penjual atau pemilik properti. Data yang digunakan sebanyak 361 data.

2. Pra-pemrosesan

Data kemudian harus diproses untuk memastikan kebersihan dan kecocokan untuk model machine learning. Ini melibatkan penghapusan nilai yang hilang, penanganan outlier, dan konversi variabel kategori menjadi format yang bisa digunakan oleh algoritma.

3. Pelatihan Model Regresi Linear

Data dibagi menjadi data pelatihan dan data uji. Model dilatih menggunakan data pelatihan untuk belajar pola dan hubungan antara fitur dan harga properti.

4. Evaluasi Model

Model dievaluasi menggunakan metrik evaluasi yang sesuai seperti Mean Squared Error (MSE), R-squared (R^2), atau Mean Absolute Error (MAE) untuk memahami seberapa baik model dapat memprediksi harga rumah.

5. Prediksi

Setelah model dioptimalkan dan dievaluasi dengan baik, ia dapat digunakan untuk melakukan prediksi harga rumah untuk data baru atau properti yang belum dilihat sebelumnya.

4. Hasil Penelitian

4.1 Pengumpulan Data

Pada tahap ini data yang digunakan ialah data yang tersedia pada laman website rumah123.com. Isi data berupa data harga rumah yang berada pada wilayah daerah bekasi yang sedang dipasarkan oleh penjual atau pemilik properti. Data terdiri dari beberapa variabel yang dijadikan variabel X atau disebut variabel dependen dan 1 variabel yang dijadikan variabel Y atau variabel independen, dan data sebanyak 361 data. Pada variabel price in rupiah angka disebut dalam jutaan

	price in rp	bedrooms	land_size_m2	building_size_m2	certificate	electricity	floors
0	2990	4	239	272	shm-sertifikatnslamlik	4400	2
1	1270	3	55	69	lbr-high-halgunanbangunan	2200	2
2	1950	3	119	131	high-halgunanbangunan	2200	2
3	3300	3	180	174	shm-sertifikatnslamlik	9500	2
4	4500	4	328	196	shm-sertifikatnslamlik	3500	2
...
366	1150	3	90	170	shm-sertifikatnslamlik	2200	2
367	1100	4	182	220	shm-sertifikatnslamlik	1300	1
388	3400	3	114	70	shm-sertifikatnslamlik	7300	1
369	2400	3	220	300	shm-sertifikatnslamlik	2200	2
380	1200	3	94	130	shm-sertifikatnslamlik	3500	2

361 rows x 7 columns

Gambar 2. Dataset harga rumah

4.2 Pra-pemrosesan

Berikut data yang sudah tampilkan statistika, mulai dari mean, nilai minimum hingga nilai maksimum harga rumah

	price in rp	bedrooms	land_size_m2	building_size_m2	electricity	floors
count	361.000000	361.000000	361.000000	361.000000	361.000000	361.000000
mean	1913.687208	2.944258	120.100268	114.797784	2877.702891	1.601350
std	1755.886421	2.071212	100.038218	92.174532	2728.871242	0.542354
min	1270.000000	1.000000	37.000000	25.000000	470.000000	1.000000
25%	692.000000	2.000000	70.000000	54.000000	1300.000000	1.000000
50%	975.000000	3.000000	90.000000	67.000000	2200.000000	1.000000
75%	1590.000000	3.000000	120.000000	140.000000	2300.000000	2.000000
max	9000.000000	35.000000	1070.000000	730.000000	47500.000000	3.000000

Gambar 3. Statistika Dataset

Dapat dilihat pada tabel, nilai minimum harga rumah sebesar 125 juta dan nilai maksimum harga rumah sebesar 9.000 juta atau 9 milyar dalam rupiah.

Selanjutnya kita mengubah nilai-nilai dalam dataset menjadi nilai-nilai yang lebih sesuai untuk pemrosesan oleh model regresi linear, variabel yang sebelumnya bertipe object diubah menjadi bilangan integer, berikut hasil dataset yang sudah ditransform data

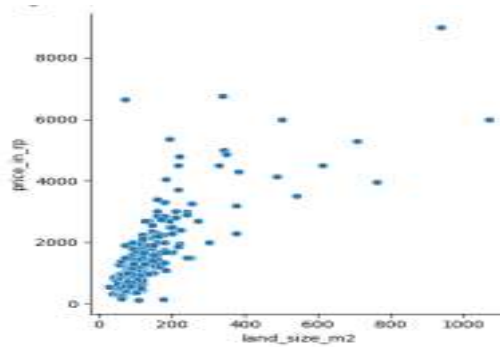
	price_in_rp	bedrooms	land_size_m2	building_size_m2	certificate	electricity	floors
0	2990	4	239	272	1	6	2
1	1270	3	55	69	0	3	2
2	1950	3	119	131	0	3	2
3	3300	3	180	174	1	5	2
4	4500	4	328	196	1	5	2

Gambar 4. Dataset yang telah ditransform

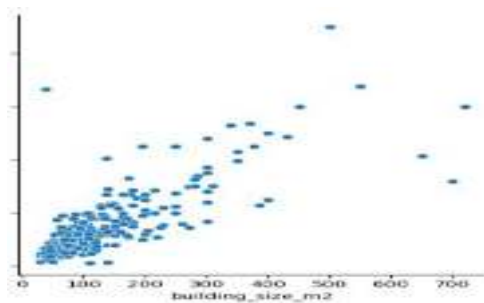
4.3. Pelatihan Model Regresi Linear

Data training dan data testing dibagi dengan rasio 80:20 dengan rincian data sebanyak 288 data training dan sebanyak 73 data pada data testing.

berikut visualisasi data hubungan variabel luas tanah, luas bangunan terhadap harga jual rumah.



Gambar 5. Hubungan variabel luas tanah terhadap harga jual rumah



Gambar 5. Hubungan variabel luas bangunan terhadap harga jual rumah

pencarian nilai slope/koefisien (m) dan intercept (b) dan Melakukan kedalaman dataframe agar lebih rapi

	features	coef_value
0	bedrooms	73.206577
1	land_size_m2	3.110501
2	building_size_m2	3.462470
3	certificate	-79.134022
4	electricity	138.119112
5	floors	252.693838

Gambar 7. Dataframe koefisien
 Mengetahui nilai korelasi dari independent variable dan dependent variable, dapat dilihat dari gambar dibawah ini hubungan variabel luas bangunan terhadap harga rumah sangat besar mendekati titik 1 sebesar 0,80, hal ini menunjukkan bahwa luas harga rumah sangat berpengaruh terhadap luas bangunan, sedangkan variabel sertifikat bernilai 0, hal ini menunjukkan bahwa variabel tersebut tidak mempengaruhi harga rumah.

	price_in_rp	bedrooms	land_size_m2	building_size_m2	certificate	electricity	floors
price_in_rp	1.00	0.50	0.78	0.80	0.00	0.71	0.42
bedrooms	0.50	1.00	0.61	0.53	0.05	0.55	0.18
land_size_m2	0.78	0.61	1.00	0.90	0.03	0.64	0.19
building_size_m2	0.80	0.53	0.90	1.00	0.02	0.70	0.44
certificate	0.00	0.05	0.03	0.02	1.00	0.06	-0.01
electricity	0.71	0.55	0.64	0.70	0.06	1.00	0.44
floors	0.42	0.18	0.19	0.44	-0.01	0.44	1.00

Gambar 8. Nilai korelasi dari independent variable dan dependent variable

Setelah dilakukan prediksi dengan algoritma regresi linear didapatkan nilai sebesar 0,8641754

4.4. Evaluasi Model

Model dievaluasi menggunakan metrik evaluasi yang sesuai dengan menguji akurasi dan kinerja dari Mean Squared Error (MSE), R-squared (R^2), atau Mean Absolute Error (MAE). Didapatkan hasil sebagai berikut :

Tabel 1. Uji Akurasi dan Kinerja

No	Evaluasi	Hasil
1	Mean Squared Error (MSE)	464,87397
2	R-squared (R^2)	0,7818811
3	Mean Absolute Error (MAE)	313,00083

Karena dibutuhkan perbandingan dalam menentukan baik atau tidaknya model regresi linear, maka perbandingan dilakukan dengan melihat Mean Squared Error (MSE), R-squared (R^2), atau Mean Absolute Error (MAE).

4.5. Prediksi

Setelah model dioptimalkan dan dievaluasi dengan baik, ia dapat digunakan untuk melakukan prediksi harga rumah untuk data baru atau properti yang belum dilihat sebelumnya. Dicontohkan pada seorang pembeli baru yang menginginkan rumah dengan banyak kamar tidur 3, luas tanah 100 meter persegi, luas bangunan 70 meter persegi, memiliki sertifikat hak milik, daya listrik 2.200 VA dan dengan bangunan berlantai dasar.

```
/usr/local/lib/python3.10/dist-packa
warnings.warn(
array([939.6174856])
```

Gambar Prediksi harga rumah sesuai keinginan

Dapat dilihat dari gambar hasil prediksi pada harga rumah yang diinginkan seorang pembeli baru tersebut menghasilkan nilai harga rumah sebesar 939,6174856 bisa dibulatkan pada angka 940 juta rupiah.

5. Kesimpulan

Dalam menentukan prediksi harga rumah ditentukan oleh banyak faktor. Berdasarkan permasalahan tersebut maka setelah dilakukan penelitian dengan menggunakan metode regresi linear untuk memprediksi harga rumah sehingga dapat mempermudah transaksi jual dan beli. Data yang digunakan dalam penelitian ini adalah data sekunder yang bersumber dari website rumah123.com. Hasil penelitian ini menunjukkan nilai akurasi sebesar 86,41%. Harga rumah sangat berpengaruh kuat dengan luas bangunan dan harga rumah tidak dipengaruhi pada kriteria jenis sertifikat dikarenakan hasil yang didapat pada keterkaitan atau hubungan variabel luas bangunan sebesar 0,80 dan jenis sertifikat sebesar 0,00. Perlunya dirancang suatu aplikasi berbasis web atau aplikasi mobile yang dapat digunakan untuk memprediksi harga rumah dengan menggunakan kriteria yang diinginkan.

6. Pustaka

- [1] L. Desyanita and A. Wibowo, "Pemodelan Sistem Prediksi Kelayakan Pengajuan Kredit," *J. Ilm. Elektron. Dan Komput.*, vol. 13, no. 2, pp. 10–22, 2020.
- [2] N. Hadi and J. Benedict, "Implementasi Machine Learning Untuk Prediksi Harga Rumah Menggunakan Algoritma Random Forest," *Comput. J. Comput. Sci. Inf. Syst.*, vol. 8, no. 1, pp. 50–61, 2024.
- [3] K. Muhammad, B. Azhary, and S. Arsitektur, "Analisis Tren Kepemilikan Rumah di Kota Palembang dan Prediksi Harga Rumah memanfaatkan Machine Learning Analysis of Home Ownership Trends in Palembang City and House Price Prediction with Machine Learning," *Perenc. Wil. PPS UHO*, vol. 8, no. 2, pp. 165–173, 2023.
- [4] A. Fadilah, A. S. R. Ansori, and R. Wijaya, "PREDIKSI HARGA RUMAH DI KOTA BANDUNG BAGIAN TIMUR DENGAN MENGGUNAKAN METODE MOVING AVERAGE," in *e-Proceeding of Engineering*, 2020, vol. 7, no. 2, pp. 4902–4911.
- [5] I. W. P. Pratama, "Implementasi Support Vector Machine Dalam Prediksi Harga Rumah," *JAVOK J. Akad. Vokasi*, vol. 2, no. 2, pp. 101–113, 2023.
- [6] V. Ariyani, P. Putri, A. B. Prasetyo, and D. Eridani, "Perbandingan Kinerja Algoritme Naïve Bayes Dan K-Nearest Neighbor (Knn) Untuk Prediksi Harga Rumah," *J. Ilm. Tek. Elektro*, no. 4, 2022.
- [7] Warjiyono, A. N. Rais, I. Alfarobi, S. W. Hadi, and W. Kurniawan, "Analisa Prediksi Harga Jual Rumah Menggunakan Algoritma Random Forest Machine Learning," *JURSISTEKNI (Jurnal Sist. Inf. dan Teknol. Informasi)*, vol. 6, no. 2, pp. 416–423, 2024.
- [8] C. Haryanto, N. Rahaningsih, and F. Muhammad Basysyar, "Komparasi Algoritma Machine Learning Dalam Memprediksi Harga Rumah," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 7, no. 1, pp. 533–539, 2023.
- [9] A. F. Kurniawan, S. F. Pane, and R. M. Awangga, "Prediksi Jumlah Penjualan Rumah di Bojongsoang ditengah Pandemi Covid-19 dengan Metode ARIMA," *J. Media Inform. Budidarma*, vol. 5, no. 4, p. 1479, 2021.
- [10] I. Maula, L. U. Hasanah, and A. Tholib, "Analisis Prediksi Harga Rumah Di Jabodetabek Menggunakan Multiple Linear Regression," *J. Inform.*, vol. 17, no. 3, pp. 238–245, 2021.
- [11] I. Kurniawan, N. Rahaningsih, and T. Suprati, "Implementasi Algoritma Regresi Linier Dan K-Nearest Neighbor Untuk Prediksi Harga Rumah," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 8, no. 1, pp. 1187–1193, 2024.
- [12] E. F. Rahayuningtyas, F. N. Rahayu, and Y. Azhar, "Prediksi Harga Rumah Menggunakan General Regression Neural Network," *J. Inform.*, vol. 8, no. 1, pp. 59–66, 2021.
- [13] N. K. Zainal, "PREDIKSI HARGA REAL ESTATE MENGGUNAKAN METODE REGRESI LINEAR BERBASIS MACHINE LEARNING," *J. Artif. Intell. Appl. (JAIA)*, vol. 1, no. 1, pp. 19–27, 2024.
- [14] A. Vermaysha and Nurmalitasari, "Prediksi Harga Rumah di Kabupaten Karanganyar Menggunakan Metode Regresi Linear," in *Prosiding Seminar Nasional Teknologi Informasi dan Bisnis (SENATIB)*, 2023, pp. 6–11.
- [15] U. Hayadi Umar and R. D. Putra, "Analisis Kenaikan Harga Properti Residensial Sederhana Untuk Wilayah Kelurahan Belian," *J. Civ. Eng. Plan.*, vol. 1, no. 2, pp. 190–196, 2020.
- [16] I. I. Ridho, G. Mahalisa, D. R. Sari, and I. Fikri, "METODE NEURAL NETWORK UNTUK PENENTUAN AKURASI PREDIKSI HARGA RUMAH Ihda," *Technologia*, vol. 13, no. 1, pp. 56–58, 2022.