

Prediksi Harga Rumah Menggunakan Metode Regresi Lasso

Kemas Alfin Yazi Bara Mentari¹, Anggara Falih Hartono², Rhaeyvan Phasca Sanggam Sinaga³,
Anggraini Puspita Sari⁴

UPN 'Veteran' Jawa Timur^{1,2,3,4}

E-mail: 22081010143@student.upnjatim.ac.id¹, 22081010122@student.upnjatim.ac.id²,
22081010192@student.upnjatim.ac.id³

Corresponding Author E-mail: anggraini.puspita.if@upnjatim.ac.id⁴

ABSTRACT

A house is one of the basic human needs. This research aims to predict house prices using the Lasso regression method. With the increasing property market and house prices continuing to soar, accurate price predictions are important for investors, developers, and home buyers. This study uses a dataset of houses in the Greater Jakarta area, with features such as land area, building area, number of bedrooms, bathrooms, and garages. The methodology used includes data pre-processing, feature standardization, division of training and test data, and application of the Lasso regression model with hyperparameter optimization using GridSearchCV. Prediction results are visualized through scatter plots comparing actual values with predictions, as well as residual plots to assess model performance. This research provides insights into the factors that influence house prices and provides predictive tools that can assist in decision-making in the property market. This approach is expected to improve understanding of house price dynamics and facilitate better investment strategies in the real estate sector.

Kata Kunci

Prediksi, Harga Rumah,
Regresi Lasso, Machine
Learning, Pasar Properti,
Jabodetabek

Keywords

*Prediction, House Price,
Lasso Regression, Machine
Learning, Property Market,
Jabodetabek*

ABSTRAK

Rumah adalah salah satu kebutuhan pokok manusia. Penelitian ini bertujuan untuk memprediksi harga rumah menggunakan metode regresi Lasso. Dengan meningkatnya pasar properti dan harga rumah yang terus melambung, prediksi harga yang akurat menjadi penting bagi investor, pengembang, dan pembeli rumah. Studi ini menggunakan dataset rumah di wilayah Jabodetabek, dengan fitur-fitur seperti luas tanah, luas bangunan, jumlah kamar tidur, kamar mandi, dan garasi. Metodologi yang digunakan meliputi pra-pemrosesan data, standarisasi fitur, pembagian data latih dan uji, serta penerapan model regresi Lasso dengan optimasi hyperparameter menggunakan GridSearchCV. Hasil prediksi divisualisasikan melalui scatter plot yang membandingkan nilai aktual dengan prediksi, serta plot residual untuk menilai kinerja model. Penelitian ini memberikan wawasan tentang faktor-faktor yang mempengaruhi harga rumah dan menyediakan alat prediktif yang dapat membantu dalam pengambilan keputusan di pasar properti. Pendekatan ini diharapkan dapat meningkatkan pemahaman tentang dinamika harga rumah dan memfasilitasi strategi investasi yang lebih baik di sektor real estate.

PENDAHULUAN

Rumah adalah salah satu komponen kehidupan yang penting karena dapat memberikan perlindungan, keamanan dan kenyamanan, serta sebagai tempat untuk beristirahat [1][2]. Rumah adalah bangunan yang melindungi kita dari panas dan hujan [3][4]. Saat ini, semua orang ingin memiliki rumah.

Studi ini akan memprediksi harga properti yang cocok untuk saat ini karena pasar properti akan terus berkembang dan harganya akan terus melambung tinggi. Umumnya, calon pembeli memiliki beberapa kriteria untuk memutuskan melakukan pembelian rumah, baik dari segi kondisi bangunan, luas tanah, luas bangunan, serta fasilitas [2][5][6].

Penelitian ini juga melihat rumah-rumah di Jabodetabek. Prediksi harga rumah yang akurat membantu investor dan pengembang properti serta pembeli rumah yang ingin menghemat uang

mereka. Dengan mengetahui cara memprediksi harga rumah dapat membantu membuat pilihan yang lebih baik dan strategis di pasar properti[7][8][9].

TINJAUAN PUSTAKA

Regresi Lasso

Regresi LASSO pertama kali diperkenalkan oleh Robert Tibshirani (1996). Sesuai namanya regresi LASSO merupakan metode regresi berganda yang digunakan untuk shrinkage yaitu menyusutkan koefisien taksiran mendekati angka nol dan selection operator yaitu menyeleksi variabel-variabel independen sehingga menghasilkan model dengan variabel terbaik[7][10][11][12]. Selain itu, regresi LASSO juga digunakan untuk data yang kontinu dan memerlukan variabel independen yang berdistribusi normal baku [7][13][14][15].

HASIL

Metodologi

Lebih dari 1000 data rumah yang akan dijual dikumpulkan untuk penelitian ini dari rumah-rumah di Jabodetabek. Selain itu, studi ini melakukan beberapa prosedur prediksi harga rumah [16][17]:

1. Memuat Data

```
data = pd.read_excel("datasrumah.xlsx")
```

2. Preproses Data

```
data = data.drop(['house_name'], axis=1)  
data['price'] = data['price'] / 1000000000
```

Menghapus kolom nama karena tidak relevan dengan perhitungan prediksi harga rumah dan mengubah harga ke satuan miliar.

3. Memisahkan fitur dan label

```
X = data[['land_area', 'building_area', 'bedroom_count', 'bathroom_count', 'number_of_garages']]  
y = data['price']
```

4. Standarisasi fitur

```
scaler = StandardScaler()  
X_scaled = scaler.fit_transform(X)  
menormalisasi data untuk digunakan dalam model
```

5. Membagi data

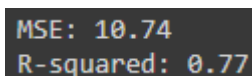
```
X_train, X_test, y_train, y_test = train_test_split(X_scaled, y, test_size=0.2, random_state=42)
```

6. Melatih Model dan Prediksi Nilai Aktual

```
grid_search = GridSearchCV(lasso, param_grid, scoring='r2', cv=5)  
grid_search.fit(X_train, y_train)  
best_lasso = grid_search.best_estimator_  
best_lasso.fit(X_train, y_train)  
y_pred = best_lasso.predict(X_test)
```

Hasil Penelitian

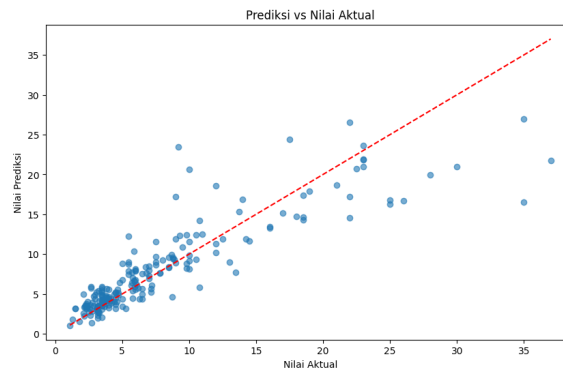
Dari penelitian yang telah dilakukan, hasil yang didapatkan adalah:



```
MSE: 10.74  
R-squared: 0.77
```

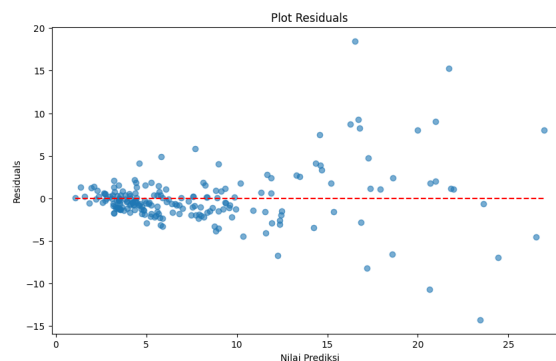
Gambar 1. Hasil Nilai

Pada gambar 1 didapatkan hasil MSE nya 10.74, MSE digunakan untuk mengetahui seberapa besar kesalahan prediksi oleh model program, caranya dengan menghitung rata-rata dari kuadrat perbedaan sehingga mse dapat menunjukkan seberapa besar kesalahan yang dilakukan oleh program. Kemudian hasil dari R-squared nya 0.77 yang artinya model regresi Lasso dapat menjelaskan kurang lebih 77% variasi harga rumah berdasarkan variabel input (luas tanah, luas bangunan, jumlah kamar tidur, jumlah kamar mandi, jumlah garasi).



Gambar 2. Grafik Nilai

Pada gambar 2 adalah bentuk visualisasi data aktual dibandingkan dengan data prediksi menggunakan scatter plot. semakin dekat titik dengan garis diagonal maka semakin akurat juga nilai prediksi dengan nilai aktualnya.



Gambar 3. Plot Residuals

Gambar 3 merupakan grafik plot residuals, untuk melihat pola kesalahan dan bias sistematis dalam prediksi model. titik menggambarkan residuals yang berarti apakah titik residuals berada dekat garis 0, hal ini menunjukkan kinerja program selama menghitung nilai prediksi. semakin merata dan dekat garis nol berarti program berjalan dengan konsisten

KESIMPULAN

Studi tersebut menunjukkan bahwa metode regresi Lasso efektif dalam memprediksi harga rumah di Jabodetabek. Model yang dikembangkan menghasilkan prediksi harga yang akurat dengan melihat aspek seperti luas tanah, luas bangunan, jumlah kamar tidur, kamar mandi, dan garasi.

Penggunaan teknik pra-pemrosesan data, standarisasi fitur, dan optimasi hyperparameter melalui GridSearchCV meningkatkan performa model. Visualisasi hasil melalui plot scatter dan residual menunjukkan kinerja model dan kesesuaian prediksi dengan nilai aktual.

Studi ini meningkatkan pengetahuan kita tentang faktor-faktor yang mempengaruhi harga rumah dan menawarkan alat prediktif yang membantu pengembang, investor, dan calon pembeli rumah membuat keputusan di pasar properti.

Meskipun model ini menunjukkan hasil yang menjanjikan, masih ada ruang untuk penyempurnaan lebih lanjut, terutama mengingat perubahan pasar properti yang konstan. Penelitian ini membuka pintu untuk pengembangan lebih lanjut dalam aplikasi machine learning di industri real estate. Untuk meningkatkan akurasi prediksi di masa depan, mungkin menambahkan variabel tambahan atau mengeksplorasi metode machine learning lainnya.

DAFTAR PUSTAKA

- [1]. Seng Jia Xin, Kamil Khalid.(2018). “Modelling House Price Using Ridge Regression and Lasso Regression “. *International Journal of Engineering & Technology*, 7 (4.30), pp.498-501.
- [2]. Linda Uswatun Hasana, Inayatul Maul’ , Abu Tholi. (2023). “ANALISIS PREDIKSI HARGA RUMAH DI JABODETABEK MENGGUNAKAN MULTIPLE LINEAR REGRESSION”. *Jurnal Informatika Kaputama (JIK)*, Vol. 7 No.2.
- [3]. Rafif Fauzan Almahd , Windha Mega Pradnya D. (2024). “Prediksi Harga Rumah Di Kabupaten Bantul Menggunakan Algoritma Support Vector Regression”. *Jurnal Teknik Informatika dan Sistem Informasi*, Vol. 11, No. 2, Hal. 152-165.
- [4]. Favorisen Rosyking Lumbanraja, Reza Aji Saputra, Kurnia Muludi, Astria Hijriani dan Akmal Junaidi. (2021). “Implementasi Support Vector Machine dalam Memprediksi Harga Rumah pada Perumahan di Kota Bandar Lampung”. *Ilmu Komputer Unila Publishing Network all right reserve*, Vol 2 No. 3, pp. 327-335.
- [5]. Muhammad Arik Verdiansyah* , Suwanda. (2023). “Penerapan Metode Regresi Komponen Utama Kernel untuk Prediksi Harga Rumah”. *Bandung Conference Series: Statistics*, hal. 663-661.
- [6]. Guntur Wahyu Kusuma, Ika Yuni Wulansari. (2019). “ANALISIS KEMISKINAN DAN KERENTANAN KEMISKINAN DENGAN REGRESI RIDGE, LASSO, DAN ELASTIC-NET DI PROVINSI JAWA TENGAH TAHUN 2017”. *Seminar Nasional Official Statistics 2019: Pengembangan Official Statistics dalam mendukung Implementasi SDG’s*, hal. 503-513.
- [7]. Muhammad Robbani, Fitriani Agustiani, Nar Herrhyanto. (2019). “REGRESI LEAST ABSOLUTE SHRINKAGE AND SELECTION OPERATOR (LASSO) PADA KASUS INFLASI DI INDONESIA TAHUN 2014-2017”. *Jurnal Eureka Matika*, Vol. 7, No. 2, hal. 1-16.
- [8]. Imi Sartika, Naomi Nussyana Debataraaja, Nurfitri Imro’ah. (2020). “ANALISIS REGRESI DENGAN METODE LEAST ABSOLUTE SHRINKAGE AND SELECTION OPERATOR (LASSO) DALAM MENGATASI MULTIKOLINEARITAS”. *Buletin Ilmiah Math. Stat. dan Terapannya (Bimaster)*, Volume 09, No. 1, hal 31 – 38.

- [9]. Kgs Muhammad Benyamin Azhary. (2023). “Analisis Tren Kepemilikan Rumah di Kota Palembang dan Prediksi Harga Rumah memanfaatkan Machine Learning”. JURNAL PERENCANAAN WILAYAH PPS UHO, Vol.8, No.2, hal. 165-173.
- [10]. Umami Athiyah , Arnelka Hananta , Taufik Maulidi , Vico Meylana Eka Putra , Theo Felix Harianto Purba, Elisabeth Angeline Wilhelmina Bakowatu. (2021). “Sistem Pendukung Keputusan Prediksi Harga Rumah Kost untuk Mahasiswa IT Telkom Purwokerto Menggunakan Metode Fuzzy Tsukamoto Berbasis Web”. Journal of Dinda, Vol. 1 No. 2, hal. 77 - 81.
- [11]. Ihda Innar Ridho, Galih Mahalisa, Dwi Retno Sari, Ihsanul Fikri. (2022). “METODE NEURAL NETWORK UNTUK PENENTUAN AKURASI PREDIKSI HARGA RUMAH”. Technologia, Vol 13, No. 1, hal. 56-58.
- [12]. Evita Fitri. (2023). “Analisis Perbandingan Metode Regresi Linier, Random Forest Regression dan Gradient Boosted Trees Regression Method untuk Prediksi Harga Rumah”. JOURNAL OF APPLIED COMPUTER SCIENCE AND TECHNOLOGY (JACOST), Vol. 4 No. 1, hal. 58 – 64.
- [13]. Cep Haryanto, Nining Rahaningsih, Fadhil Muhammad Basysyar. (2023). “KOMPARASI ALGORITMA MACHINE LEARNING DALAM MEMPREDIKSI HARGA RUMAH”. Jurnal Mahasiswa Teknik Informatika, Vol. 7 No. 1, hal. 533-539.
- [14]. Muhammad Labib Mu'tashim, Sekar Ayu Damayanti, Hanan Nadia Zaki, Toni Muhayat, Rio Wirawan. (2021). “Analisis Prediksi Harga Rumah Sesuai Spesifikasi Menggunakan Multiple Linear Regression”. JURNAL INFORMATIK Edisi ke-17, Nomor 3, hal. 238-245.
- [15]. Evi Febrion Rahayuningtyas , Feny Novia Rahayu , Yufis Azhar. (2021). “Prediksi Harga Rumah Menggunakan General Regression Neural Network”. JURNAL INFORMATIKA, Vol.8 No.1 April 2021, Halaman 59~66.
- [16]. Andi Saiful, Septi Andryana, Aris Gunaryati. (2021). “Prediksi Harga Rumah Menggunakan Web Scrapping Dan Machine Learning Dengan Algoritma Linear Regression”. Jurnal Teknik Informatika dan Sistem Informasi Vol. 8, No. 1, Hal. 41-50.
- [17]. Anggraini Puspita Sari, Andreas Nugroho Sihananto, Dwi Arman Prasetya, Muhammad Muharrom Al Haromainy. (2022). “Pengelompokan Tingkat Penyebaran Covid 19 Pada Kabupaten Jombang dengan Menggunakan Algoritma K-Means ”. Jurnal Teknologi Informasi dan Komunikasi (SCAN), Vol. 17, No. 3, Hal. 8-12.