

Prediksi Harga Rumah di Kabupaten Karanganyar Menggunakan Metode Regresi Linear

Ashary Vermaysha^{1*}, Nurmalitasari²

^{1,2}Sistem Informasi

Universitas Duta Bangsa Surakarta

^{1*}vermaysha@gmail.com, ²nurmalitasari@udb.ac.id

Abstrak—Penelitian ini bertujuan untuk menerapkan metode regresi linear dalam prediksi harga rumah di Kabupaten Karanganyar, dengan menggunakan luas tanah, luas bangunan, jumlah kamar mandi, jumlah kamar tidur, dan daya listrik sebagai variabel independen. Sebuah dataset yang terdiri dari harga rumah dan data independen tersebut digunakan dalam analisis. Hasil penelitian menunjukkan bahwa prediksi harga rumah yang dihasilkan memiliki tingkat akurasi yang relatif kurang, dengan kesalahan sebesar 651614542,27. Untuk meningkatkan akurasi prediksi, disarankan tiga langkah. Pertama, menambahkan variasi data dari beberapa website jual-beli rumah untuk memperluas sampel dan karakteristik rumah yang tercakup. Kedua, memperluas dataset dengan menambahkan variabel-variabel pendukung seperti tanggal, yang memberikan informasi tentang tren dan fluktuasi pasar. Terakhir, eksplorasi algoritma prediksi alternatif yang lebih mendukung selain regresi linear. Kesimpulan penelitian ini adalah bahwa meskipun regresi linear dapat digunakan untuk memprediksi harga rumah di Kabupaten Karanganyar, tingkat akurasinya masih perlu ditingkatkan. Oleh karena itu, rekomendasi yang diusulkan perlu diimplementasikan untuk memperbaiki model prediksi. Dengan demikian, diharapkan prediksi harga rumah di Kabupaten Karanganyar menjadi lebih akurat dan bermanfaat untuk digunakan dalam pasar properti.

Kata kunci—*prediksi, harga, rumah, regresi linear, rmse*

Abstract—This research aims to apply linear regression method in predicting house prices in Karanganyar Regency, using land area, building area, number of bathrooms, number of bedrooms, and electricity capacity as independent variables. A dataset consisting of house prices and the independent data was used in the analysis. The research results show that the predicted house prices have a relatively low level of accuracy, with an error of 651614542,27. To improve the prediction accuracy, three steps are recommended. First, adding data variations from several real estate websites to expand the sample and covered house characteristics. Second, expanding the dataset by adding supporting variables such as dates, which provide information about market trends and fluctuations. Finally, exploring alternative prediction algorithms that are more supportive than linear regression. The conclusion of this research is that although linear regression can be used to predict house prices in Karanganyar Regency, the accuracy level still needs to be improved. Therefore, the proposed recommendations need to be implemented to improve the prediction model. Thus, it is expected that house price predictions in Karanganyar Regency will become more accurate and useful for use in the property market.

Keywords—*prediction, price, house, linear regression, rmse*.

I. PENDAHULUAN

Perkembangan sektor properti di Indonesia telah mengalami peningkatan yang signifikan dalam beberapa tahun terakhir. Salah satu faktor penting dalam sektor properti adalah harga rumah, yang menjadi pertimbangan utama bagi para pembeli. Kabupaten Karanganyar, yang terletak di Provinsi Jawa Tengah, mengalami perkembangan yang pesat dalam hal pembangunan perumahan. Pertumbuhan ekonomi, urbanisasi, dan peningkatan infrastruktur di Kabupaten Karanganyar telah menyebabkan peningkatan permintaan rumah yang tinggi di daerah tersebut. Oleh karena itu, penting untuk memiliki alat prediksi yang akurat untuk memperkirakan harga rumah di Kabupaten Karanganyar.

Dalam hal ini, maka penggunaan regresi linear dapat menjadi metode yang efektif untuk

memprediksi harga rumah. Regresi linear adalah metode statistik yang digunakan untuk memodelkan hubungan antara variabel dependen (harga rumah) dengan satu atau lebih variabel independen (faktor-faktor yang mempengaruhi harga rumah). Variabel yang memiliki pengaruh signifikan terhadap harga rumah di Kabupaten Karanganyar antara lain luas tanah, luas bangunan, jumlah kamar tidur, jumlah kamar mandi, dan daya listrik. Dengan menggunakan metode regresi linear, kita dapat mengidentifikasi pola dan hubungan yang ada antara variabel-variabel ini dengan harga rumah di Kabupaten Karanganyar.

Penelitian ini bertujuan untuk menerapkan model prediksi harga rumah di Kabupaten Karanganyar menggunakan regresi linear dengan mempertimbangkan variabel-variabel yang

mempengaruhi, yaitu luas tanah, luas bangunan, jumlah kamar tidur, jumlah kamar mandi, dan daya listrik. Dengan memiliki model prediksi yang akurat, diharapkan dapat memperkirakan harga rumah di Kabupaten Karanganyar berdasarkan faktor-faktor tersebut. Hal ini akan memberikan manfaat bagi berbagai pihak, termasuk pengembang properti dalam perencanaan proyek, calon pembeli rumah dalam pengambilan keputusan, serta pemerintah dalam merumuskan kebijakan perumahan yang lebih efektif. Selain itu, penelitian ini juga diharapkan dapat memberikan wawasan baru dalam memahami faktor-faktor yang mempengaruhi harga rumah di daerah tersebut.

II. METODE PENELITIAN

1.1. Prediksi

Prediksi merupakan suatu kegiatan untuk memperkirakan apa yang akan terjadi pada masa yang akan datang [1] dengan memperhatikan data maupun informasi dimasa lampau atau masa sekarang baik secara matematik maupun statistik [2] yang bertujuan untuk mengetahui, melihat dan memperkirakan prospek ekonomi atau kegiatan usaha. [3]

Menurut [4] dalam perancangan suatu metode prediksi diperlukan tiga tahapan yang harus dilalui antara lain:

1. Melakukan analisa data masa lampau yang bertujuan untuk mendapatkan gambaran maupun pola dari data yang bersangkutan.
2. Memilih metode prediksi yang akan digunakan. Terdapat bermacam-macam metode prediksi yang tersedia dengan keperluannya masing-masing, setiap metode dapat mempengaruhi hasil dari prediksi yang dilakukan. Oleh karena itu tidak ada metode yang pasti cocok untuk semua jenis data.
3. Proses transformasi data dari data masa lampau dengan menggunakan metode yang dipilih, apabila diperlukan adanya transformasi data maka akan dilakukan perubahan sesuai kebutuhannya.

1.2. Regresi Linear

Regresi linear merupakan salah satu metode statistika yang memberikan hasil berupa prediksi dengan melakukan pengembangan hubungan sistematis antara satu atau lebih variabel independen dengan satu variabel dependen [5]. Tujuan dari regresi linear yaitu untuk menemukan persamaan garis lurus yang paling cocok untuk menggambarkan hubungan antara variabel-variabel tersebut.

Regresi linear bergantung pada beberapa asumsi penting, termasuk asumsi tentang hubungan linier antara variabel, distribusi normalitas dari kesalahan (selisih antara nilai sebenarnya dan nilai yang diprediksi), homoskedastisitas (varians kesalahan konstan), dan independensi dari kesalahan.

Metode regresi linear sering digunakan dalam berbagai bidang, termasuk ekonomi, ilmu sosial, ilmu lingkungan, dan banyak lagi. Dengan memahami hubungan antara variabel-variabel yang relevan, regresi linear dapat memberikan pemahaman yang lebih dalam tentang faktor-faktor yang mempengaruhi suatu fenomena dan dapat digunakan untuk membuat prediksi yang berguna. Adapun untuk perhitungan *regresi linear* yang digunakan.

$$Y = \beta^0 + \beta^1 X^1 + \beta^2 X^2 + \dots + \beta_p X_p + \varepsilon \quad (1)$$

dengan

X_1, X_2, \dots, X_p variabel independen (variabel prediktor).
variabel dependen (variabel yang ingin diprediksi).

β_0, \dots, β_p koefisien regresi yang masing-masing menggambarkan perubahan rata-rata dalam Y ketika X_1, X_2, \dots, X_p berubah satu satuan, intercept (nilai konstanta) yang menunjukkan titik perpotongan garis regresi dengan sumbu y.

kesalahan/residual, yang merupakan selisih antara nilai sebenarnya dan nilai yang diprediksi oleh model.

Maka pernyataan diatas dapat dijadikan sebagai acuan untuk menghitung prediksi harga rumah di Kabupaten Karanganyar dengan algoritma *regresi linear* pada penelitian ini dan dapat dinyatakan dalam pernyataan

$$Y = \beta^0 + \beta^1 X^1 + \beta^2 X^2 + \beta^3 X^3 + \beta^4 X^4 + \beta^5 X^5 + \varepsilon \quad (2)$$

dengan

Y	harga yang merupakan harga rumah yang ingin diprediksi
X ₁	= luas tanah
X ₂	= luas bangunan
X ₃	= jumlah kamar mandi
X ₄	= jumlah kamar tidur
X ₅	= daya listrik
β ₀ , β ₁ , β ₂ , β ₃ , β ₄ , dan β ₅	= adalah koefisien regresi yang menggambarkan pengaruh relatif dari masing-masing variabel independen terhadap harga rumah.
ε	= kesalahan atau residual yang merupakan selisih antara nilai sebenarnya dan nilai yang diprediksi oleh model.

1.3. Pengujian Keakuratan Hasil

Keakuratan hasil prediksi merupakan aspek penting dalam analisis regresi linear. Hal ini mencerminkan sejauh mana model dapat memprediksi nilai yang mendekati nilai sebenarnya. Terdapat beberapa metrik evaluasi yang umum digunakan untuk mengukur keakuratan hasil prediksi, antara lain Mean Absolute Percentage Error (MAPE), Root Mean Squared Error (RMSE), dan Relative Mean Error (RME).

MAPE (Mean Absolute Percentage Error) adalah salah satu metrik evaluasi yang mengukur persentase kesalahan rata-rata antara nilai sebenarnya dan nilai prediksi. Rumus MAPE dapat dinyatakan sebagai berikut:

$$MAPE = \left(\frac{1}{n}\right) * \Sigma \left(\left|\frac{y_i - \bar{y}}{y_i}\right|\right) * 100 \quad (3)$$

dengan

n	= jumlah data yang digunakan dalam perhitungan
---	--

y _i	= nilai sebenarnya dari data ke-i
\bar{y}	= nilai yang diprediksi untuk data ke-i.

RMSE (Root Mean Squared Error) sering digunakan untuk mengevaluasi hasil kinerja prediksi yang digunakan [5] dengan mengukur tingkat keakuratan dari hasil prediksi. Jika nilai RMSE relatif rendah maka dapat disimpulkan bahwa bentuk perubahan nilai yang dihasilkan oleh model prediksi mendekati bentuk aslinya, sebaliknya jika nilai RMSE relatif besar maka keakuratan yang dihasilkan oleh model prediksi jauh dari bentuk aslinya. Tahap pengujian akurasi hasil prediksi menggunakan nilai RMSE (Root Mean Squared Error).

$$RMSE = \sqrt{(1/n) * \Sigma (y_i - \bar{y})^2} \quad (4)$$

dengan

n	= jumlah data yang digunakan dalam perhitungan.
y _i	= nilai sebenarnya dari data ke-i.
\bar{y}	= nilai yang diprediksi untuk data ke-i.
Σ	= menunjukkan operasi penjumlahan, yang dilakukan untuk setiap data dari i = 1 hingga n.

RME (Relative Mean Error) adalah metrik evaluasi yang mengukur rata-rata kesalahan relatif antara nilai sebenarnya dan nilai prediksi. Rumus RME dapat dinyatakan sebagai berikut:

$$RME = \left(\frac{1}{n}\right) * \Sigma \left(\frac{|y_i - \bar{y}|}{y_i}\right) \quad (5)$$

dengan

n	= jumlah data yang digunakan dalam perhitungan.
y _i	= nilai sebenarnya dari data ke-i.
\bar{y}	= nilai yang diprediksi untuk data ke-i.

Metrik evaluasi seperti MAPE, RMSE, dan RME membantu dalam mengevaluasi dan membandingkan keakuratan hasil prediksi model regresi linear. Semakin kecil nilai MAPE, RMSE, dan RME, semakin baik prediksi model tersebut.

1.4. Pengumpulan Data

Penelitian ini menggunakan pengumpulan data untuk prediksi harga rumah yang dilakukan dengan studi dokumenter pada website jual beli rumah terpercaya, ini dilakukan dikarenakan tidak semua website dapat memenuhi variabel dalam penelitian ini. Terdapat 5

variabel yang mempengaruhi harga suatu rumah yaitu luas lahan, luas bangunan tersebut, jumlah kamar tidur, jumlah kamar mandi, dan daya listrik.

Dengan melakukan web scrapping memungkinkan untuk mengumpulkan data dengan jumlah besar dan mendapatkan sumber data terkait.

1.5. Pengolahan Data

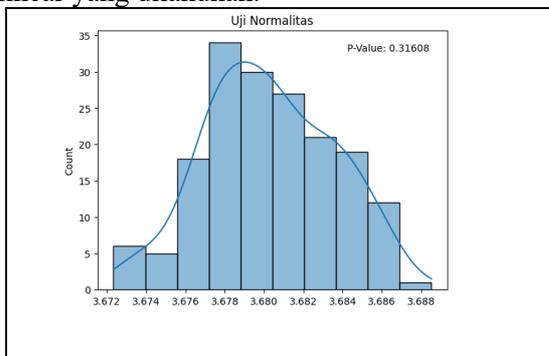
Tahap ini digunakan untuk menghasilkan nilai yang lebih baik, serta untuk menghapus data-data yang tidak valid selama proses scrapping. Data yang tidak valid termasuk juga data yang rusak maupun data yang tidak diinginkan.

III. HASIL PENELITIAN DAN PEMBAHASAN

1.1. Uji Statistik

1.1.1. Uji Normalitas

Normalitas adalah salah satu asumsi penting dalam regresi linear, yang mengasumsikan bahwa suatu variabel dependen memiliki distribusi normal. Dengan dilakukannya uji normalitas dapat membantu dalam memberikan kepercayaan kedalam hasil interpretasi dari regresi linear yang dilakukan.

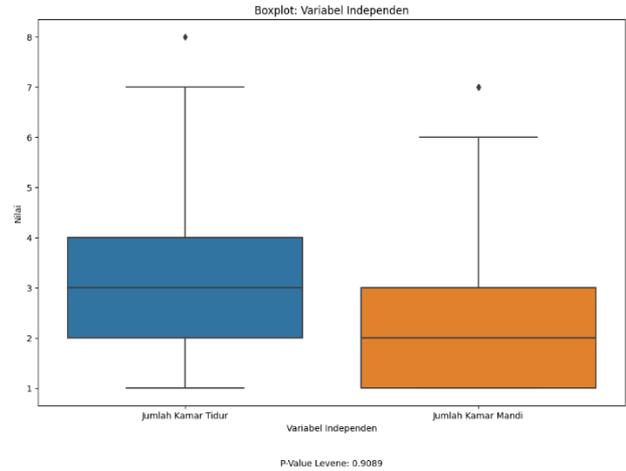


Gambar 1 Uji Normalitas

Dari gambar diatas dapat diambil kesimpulan bahwa dataset yang diberikan berdistribusi normal, dikarenakan p-value yang ditampilkan lebih dari nilai alpha yakni 0,05.

1.1.2. Uji Homogenitas

Pada regresi linear penting untuk memeriksa apakah varians dari residu (selisih antara nilai aktual dan nilai yang diprediksi oleh model) dalam model regresi. Homogenitas varians adalah salah satu asumsi yang penting dalam regresi linear yang mengasumsikan bahwa variabel dependen memiliki varians yang sama disemua tingkat variabel independen.

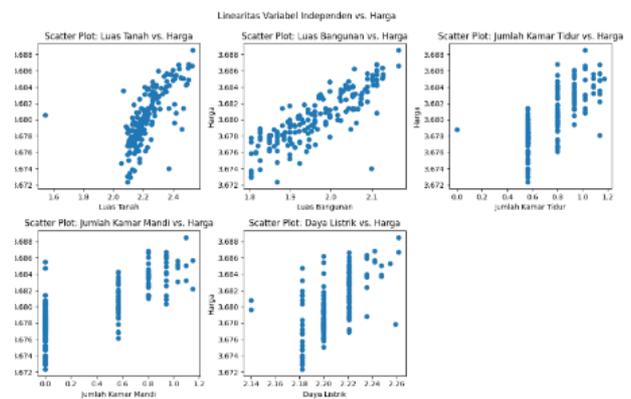


Gambar 2 Uji Homogenitas

Dari grafik diatas maka dapat disimpulkan bahwa dataset yang diberikan terbukti homogen dikarenakan p-value yang ditampilkan lebih dari 0,05 yaitu dengan nilai 0,908 maka dapat diambil kesimpulan bahwa kelompok data memiliki varians yang sama (homogen)

1.1.3. Uji Linearitas

Uji ini bertujuan untuk memverifikasi apakah hubungan antara variabel independen dan variabel dependen bersifat linier. Dalam regresi linear, diasumsikan bahwa hubungan antara variabel independen dan variabel dependen adalah linier, sehingga uji linearitas digunakan untuk menguji validitas asumsi.



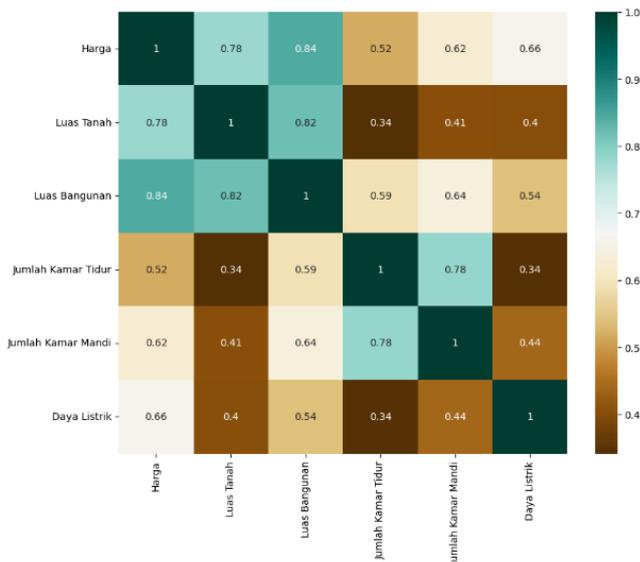
Gambar 3 Uji Linearitas

Dari gambar (3) dapat dilihat bahwa setiap variabel independen tersebar dengan membentuk pola yang cenderung membentuk pola garis lurus, dengan begitu dapat diambil kesimpulan bahwa dataset yang digunakan bersifat linear

1.1.4. Korelasi

Korelasi antar variabel adalah konsep yang digunakan dalam statistik untuk mengukur sejauh mana hubungan atau hubungan linier antar dua atau lebih variabel. Korelasi memberikan indikasi tentang arah dan kekuatan hubungan antar variabel-variabel.

Pada gambar 4 dapat dilihat tabel korelasi antar variabel dengan yang lainnya menunjukkan hasil yang paling mempengaruhi harga adalah variabel luas tanah sedangkan yang tidak mempengaruhi harga adalah jumlah kamar tidur.



Gambar 4 Korelasi antar variabel

1.2. Uji Keakuratan Hasil Prediksi

1.2.1. Koefisien Determinasi

Koefisien Determinasi adalah salah satu metrik yang digunakan untuk mengukur sejauh mana variabilitas variabel dependen dapat dijelaskan oleh variabel independen yang digunakan. Nilai koefisien determinasi yang mendekati dengan 1 menunjukkan bahwa variabel independen dalam model secara efektif menjelaskan variasi yang ada dalam variabel dependen, sebaliknya apabila nilai koefisien determinasi mendekati 0 maka menunjukkan bahwa model tidak berhasil menjelaskan variasi variabel dependen.

Koefisien determinasi (R-squared): 0.8151731662898201

Gambar 5 Nilai R-Squared

Berdasarkan gambar 5, dapat diketahui bahwa nilai R-Squared sebesar 0,815 atau 81,5% maka dapat ditarik

kesimpulan bahwa pengaruh variabel independen terhadap variabel dependen yaitu 81,5%. Sedangkan sisa 18,5% dipengaruhi oleh variabel lain yang tidak termasuk dalam model.

1.2.2. RMSE

RMSE (Root Mean Squared Error) merupakan metrik evaluasi yang mengukur seberapa akurat model regresi dalam melakukan prediksi terhadap variabel dependen.

Root Mean Squared Error: 651614542.27

Gambar 6 Nilai RMSE

Pada gambar 6 menunjukkan Root Mean Squared Error prediksi sebesar 651614542,27 yang berarti bahwa adanya kesalahan yang sangat besar antara nilai prediksi dan nilai sebenarnya.

1.2.3. Perbandingan Prediksi



Gambar 7 Perbandingan Nilai Prediksi dan Nilai Sebenarnya

Pada gambar 7 menunjukkan hasil nilai prediksi dengan nilai sebenarnya pada model regresi, dengan membandingkan kedua garis tersebut, dapat dilihat bahwa hasil prediksi cenderung kurang akurat apabila data yang akan diprediksi cukup besar.

IV. KESIMPULAN

Dari hasil penelitian penerapan prediksi harga rumah di Kabupaten Karanganyar menggunakan regresi linear ini dapat disimpulkan bahwa hasil prediksi yang diberikan relatif kurang akurat dengan error sebesar 651614542,27. Maka dari itu untuk meningkatkan akurasi prediksi harga rumah, penulis mengusulkan untuk melakukan hal-hal dibawah ini kepada penulis selanjutnya:

1. Menambahkan variasi data dari beberapa website jual-beli rumah.

2. Menambahkan variabel-variabel independen lainnya yang lebih relevan untuk memperkaya model regresi linear seperti tanggal.
3. Pemilihan variabel yang lebih baik dengan mempertimbangkan pengaruhnya terhadap variabel target.
4. Menerapkan algoritma prediksi lainnya yang lebih mendukung untuk melakukan prediksi secara lebih akurat

REFERENSI

- [1] E. Prasetyowati, B. Said, dan S. Rachmatullah, "Aplikasi Penentuan Harga Pokok Produksi Batik Madura Dengan Metode Activity Based Costing Dan Analisis Regresi Linier," *JUTI: Jurnal Ilmiah Teknologi Informasi*, vol. 16, no. 1, hlm. 48–58, 2018.
- [2] M. Hakimah, R. R. Muhima, dan A. Yustina, "Rancang Bangun Aplikasi Peramalan Persediaan Barang Dengan Metode Trend Projection," *Jurnal Simantec*, vol. 5, no. 1, 2015.
- [3] R. Gustriansyah, "Analisis metode single exponential smoothing dengan brown exponential smoothing pada studi kasus memprediksi kuantiti penjualan produk farmasidi apotek," *Semnasteknomedia online*, vol. 5, no. 1, hlm. 3–5, 2017.
- [4] T. Indarwati, T. Irawati, dan E. Rimawati, "Penggunaan Metode Linear Regression Untuk Prediksi Penjualan Smartphone," *Jurnal Teknologi Informasi Dan Komunikasi (TIKomSiN)*, vol. 6, no. 2, 2019.
- [5] A. Saiful, "Prediksi harga rumah menggunakan web scrapping dan machine learning dengan algoritma linear regression," *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, vol. 8, no. 1, hlm. 41–50, 2021.