



Implementasi Data Mining Dalam Analisis dan Prediksi Harga Rumah Wilayah Jakarta Menggunakan Regresi Linear Berganda

Aufaldo Rifqi Gunawan^{1*}, Siti Khotijah², Siwi Puji Astuti³

¹²³ Fakultas Teknik dan Ilmu Komputer, Program Studi Teknik Informatika, Universitas Indraprasta PGRI, Jakarta, Indonesia

Email: ^{1*}dimasf.14.y@gmail.com, ²sitik2805@gmail.com, ³siwiunindra2012@gmail.com

(* : corresponding author)

Abstrak—Perkembangan platform properti digital seperti Rumah123.com menghasilkan data listing rumah dalam jumlah besar, namun pemanfaatannya untuk menghasilkan model prediksi harga yang objektif dan dapat diakses publik masih terbatas. Penelitian ini bertujuan menerapkan teknik data mining dengan metode regresi linear berganda untuk menganalisis dan memprediksi harga rumah di lima wilayah administratif DKI Jakarta berdasarkan karakteristik fisik properti. Data dikumpulkan melalui teknik web scraping menggunakan Python (BeautifulSoup dan Requests) dari Rumah123.com, menghasilkan 750 data awal dengan delapan atribut, yang setelah praproses (pemeriksaan missing value dan penanganan outlier menggunakan metode Interquartile Range) tersisa 692 data valid. Analisis regresi menggunakan metode Ordinary Least Squares pada perangkat lunak SPSS menghasilkan model dengan koefisien determinasi (R^2) sebesar 0,550 dan Adjusted R^2 sebesar 0,543, serta terbukti signifikan secara simultan melalui uji F ($F = 83,185$; $\text{Sig.} < 0,001$). Uji asumsi klasik menunjukkan residual berdistribusi normal, tidak terjadi heteroskedastisitas, tidak ada multikolinearitas serius, dan tidak terjadi autokorelasi (Durbin-Watson = 2,017). Hasil uji t menunjukkan bahwa lokasi wilayah, luas bangunan, luas tanah, dan jumlah kamar mandi berpengaruh signifikan terhadap harga rumah, dengan luas tanah sebagai prediktor paling dominan, sedangkan jumlah kamar tidur, ketersediaan garasi, dan jumlah lantai tidak berpengaruh signifikan. Persamaan regresi yang dihasilkan selanjutnya diimplementasikan ke dalam aplikasi berbasis web menggunakan PHP dan MySQL agar dapat digunakan secara interaktif oleh masyarakat umum, calon pembeli, dan investor properti dalam menilai kewajaran harga rumah di wilayah Jakarta.

Kata Kunci: Data Mining; Regresi Linear Berganda; Prediksi Harga Rumah; Web Scraping; DKI Jakarta.

Abstract—The growth of digital property platforms such as Rumah123.com has produced a large volume of house-listing data, yet this data has rarely been used to build an objective and publicly accessible price-prediction model. This study applies a data mining approach using multiple linear regression to analyze and predict house prices in five administrative regions of DKI Jakarta based on property characteristics. Data were collected through web scraping using Python (BeautifulSoup and Requests) from Rumah123.com, yielding 750 initial records with eight attributes; after preprocessing (missing-value inspection and outlier handling using the Interquartile Range method), 692 valid records remained. Regression analysis using the Ordinary Least Squares method in SPSS produced a model with a coefficient of determination (R^2) of 0.550 and an Adjusted R^2 of 0.543, which was statistically significant based on the F-test ($F = 83.185$; $\text{Sig.} < 0.001$). The classical assumption tests indicated normally distributed residuals, no heteroscedasticity, no serious multicollinearity, and no autocorrelation (Durbin-Watson = 2.017). The t-test results show that regional location, building area, land area, and number of bathrooms significantly affect house prices, with land area as the most dominant predictor, while the number of bedrooms, garage availability, and number of floors were not significant. The resulting regression equation was subsequently implemented into a PHP- and MySQL-based web application, enabling the public, prospective buyers, and property investors to interactively assess the fairness of house prices in the Jakarta area.

Keywords: Data Mining; Multiple Linear Regression; House Price Prediction; Web Scraping; DKI Jakarta.

1. PENDAHULUAN

Perkembangan teknologi informasi yang pesat telah mendorong transformasi besar di berbagai sektor, termasuk sektor properti dan perumahan. Kemajuan kecerdasan buatan dan pembelajaran mesin memunculkan pendekatan penambangan data (data mining) yang memungkinkan ekstraksi pola bermakna dari kumpulan data berskala besar, sehingga mendukung pengambilan keputusan yang lebih akurat dan berbasis data. Salah satu sumber data yang potensial untuk dimanfaatkan adalah platform marketplace properti daring, seperti Rumah123.com, yang menyediakan ribuan listing properti residensial di berbagai wilayah termasuk DKI Jakarta, lengkap dengan atribut harga, lokasi, luas tanah, luas bangunan, jumlah kamar, dan fasilitas lainnya.



Pasar properti, khususnya harga rumah, merupakan indikator ekonomi yang dinamis dan dipengaruhi oleh berbagai faktor multivariat. Berdasarkan Indeks Harga Properti Perumahan yang diterbitkan Badan Pusat Statistik, harga rumah di Indonesia naik 2,76% pada Maret 2024 dibandingkan periode yang sama tahun sebelumnya, dan secara kumulatif meningkat 10,90% sejak 2019. Rumah123 Flash Report edisi Februari 2024 juga mencatat pertumbuhan tahunan harga rumah di wilayah Jakarta sebesar 2,7% pada Januari 2024. Meskipun demikian, ribuan data listing properti pada platform seperti Rumah123.com belum dimanfaatkan secara optimal untuk menghasilkan model prediksi harga yang dapat diakses oleh masyarakat umum, sehingga calon pembeli maupun investor kesulitan menilai kewajaran harga properti yang ditawarkan.

Beberapa penelitian terdahulu telah mengkaji prediksi harga rumah menggunakan regresi linear berganda. Mu'tashim dkk. (2021) menganalisis prediksi harga rumah sesuai spesifikasi menggunakan Multiple Linear Regression dan memperoleh akurasi model sebesar 66% ($R^2 = 0,66$), dengan luas bangunan dan luas tanah sebagai faktor paling dominan. Maula dkk. (2023) melakukan analisis prediksi harga rumah di wilayah Jabodetabek dan menemukan bahwa luas bangunan, luas tanah, dan lokasi berpengaruh signifikan terhadap harga. Hakim dkk. (2025) juga menunjukkan bahwa pendekatan pembelajaran mesin berbasis regresi linear cukup efektif untuk estimasi harga rumah. Namun, penelitian-penelitian tersebut umumnya masih terbatas pada cakupan wilayah yang sempit (satu kota atau kecamatan) dan belum mengintegrasikan mekanisme pengumpulan data otomatis dari platform properti daring dengan pemodelan statistik dan penyediaan antarmuka digital yang dapat langsung diakses masyarakat.

Berdasarkan kesenjangan tersebut, penelitian ini mengusulkan penerapan data mining dengan regresi linear berganda untuk membangun model prediksi harga rumah yang mencakup seluruh lima wilayah administratif DKI Jakarta, yaitu Jakarta Barat, Jakarta Pusat, Jakarta Selatan, Jakarta Timur, dan Jakarta Utara. Data dikumpulkan melalui teknik web scraping dari Rumah123.com dan mencakup tujuh variabel prediktor, yaitu luas tanah, luas bangunan, jumlah kamar tidur, jumlah kamar mandi, ketersediaan garasi, jumlah lantai, dan wilayah, dengan harga rumah sebagai variabel dependen. Berbeda dengan penelitian sebelumnya yang bersifat statis, penelitian ini mengintegrasikan seluruh tahapan mulai dari pengumpulan data otomatis, analisis statistik menggunakan SPSS, hingga implementasi model ke dalam aplikasi berbasis web yang dapat digunakan secara interaktif oleh masyarakat umum, calon pembeli, investor, agen, maupun pemangku kebijakan.

Penelitian ini bertujuan untuk: (1) mengumpulkan data properti residensial di lima wilayah DKI Jakarta melalui teknik web scraping; (2) menghasilkan persamaan model regresi linear berganda untuk memprediksi harga rumah; (3) mengidentifikasi variabel yang berpengaruh signifikan terhadap harga rumah; (4) merancang dan membangun aplikasi prediksi harga berbasis web menggunakan PHP dan MySQL; dan (5) mengevaluasi tingkat akurasi model yang dihasilkan. Hasil penelitian diharapkan dapat memberikan estimasi harga yang objektif, transparan, dan ilmiah, serta menjadi rujukan bagi penelitian sejenis maupun pengembangan sistem prediksi harga properti berbasis data di kota-kota metropolitan lain di Indonesia.

2. METODE PENELITIAN

Penelitian ini menggunakan metode deskriptif kuantitatif untuk memberikan gambaran sistematis mengenai pola hubungan antara karakteristik properti dan harga rumah di wilayah DKI Jakarta berdasarkan data numerik. Data yang digunakan merupakan data sekunder yang dikumpulkan dari situs Rumah123.com, salah satu platform marketplace properti terbesar di Indonesia, melalui teknik web scraping menggunakan bahasa pemrograman Python beserta pustaka BeautifulSoup dan Requests. Proses pengumpulan data menghasilkan 750 data properti dari lima wilayah administratif DKI Jakarta (Jakarta Barat, Jakarta Pusat, Jakarta Selatan, Jakarta Timur, dan Jakarta Utara), yang disimpan dalam format Microsoft Excel (.xlsx) dengan delapan atribut, yaitu lokasi, luas bangunan, luas tanah, jumlah kamar tidur, jumlah kamar mandi, ketersediaan garasi, jumlah lantai, dan harga.

Penelitian dilaksanakan melalui delapan tahap yang saling berkesinambungan, yaitu: (1) identifikasi masalah, (2) studi pustaka, (3) pengumpulan data, (4) praproses data, (5) analisis data, (6) pemodelan regresi linear berganda, (7) implementasi perangkat lunak, dan (8) penarikan kesimpulan. Tahap praproses data meliputi konversi tipe data (misalnya mengubah teks "Rp 6,38



Miliar” menjadi nilai numerik 6,38), encoding variabel kategorikal lokasi melalui label encoding dan one-hot encoding, pemeriksaan missing value, serta deteksi dan penanganan pencilon (outlier) menggunakan metode Interquartile Range (IQR) dengan batas bawah $Q1 - 1,5 \times IQR$ dan batas atas $Q3 + 1,5 \times IQR$ pada kolom luas bangunan, luas tanah, dan harga. Baris data yang memiliki nilai di luar rentang tersebut dihapus dari dataset.

Analisis data menggunakan metode regresi linear berganda dengan estimasi Ordinary Least Squares (OLS) yang diproses menggunakan perangkat lunak SPSS. Model regresi dinyatakan dalam persamaan umum $\hat{Y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \varepsilon$, dengan $\hat{Y}$ adalah harga rumah yang diprediksi, β_0 adalah konstanta, β_1 hingga β_n adalah koefisien regresi, X_1 hingga X_n adalah variabel independen, dan ε adalah residual. Sebelum interpretasi model dilakukan, empat uji asumsi klasik diterapkan untuk memastikan estimasi OLS bersifat Best Linear Unbiased Estimator (BLUE), yaitu uji normalitas residual (Kolmogorov-Smirnov dan inspeksi visual histogram serta Normal P-P Plot), uji multikolinearitas menggunakan nilai Variance Inflation Factor (VIF, dengan ambang batas $VIF < 10$ dan indikasi moderat pada $VIF 5-10$), uji heteroskedastisitas menggunakan diagram sebar residual, serta uji autokorelasi menggunakan statistik Durbin-Watson (nilai ideal mendekati 2, pada rentang 1,65–2,35).

Kelayakan model dievaluasi melalui tiga indikator utama, yaitu koefisien determinasi (R^2 dan Adjusted R^2) untuk mengukur proporsi variasi harga rumah yang dapat dijelaskan model, uji F (simultan) untuk menguji pengaruh seluruh variabel independen secara bersama-sama, dan uji t (parsial) untuk menguji signifikansi masing-masing variabel independen. Tahap akhir penelitian adalah menerjemahkan persamaan regresi yang diperoleh ke dalam aplikasi berbasis web menggunakan bahasa pemrograman PHP untuk logika aplikasi dan antarmuka pengguna, serta basis data MySQL untuk menyimpan data properti dan hasil prediksi, sehingga pengguna dapat memasukkan spesifikasi rumah melalui antarmuka web dan memperoleh estimasi harga secara otomatis dan interaktif.

3. ANALISA DAN PEMBAHASAN

3.1 Gambaran Dataset dan Praproses Data

Penelitian ini menggunakan dataset harga rumah di wilayah DKI Jakarta yang terdiri atas 750 data properti dengan delapan atribut, sebagaimana disajikan pada Tabel 1.

Tabel 1. Atribut Dataset

No.	Nama Kolom	Tipe Data	Keterangan
1	Lokasi	String/Object	Nama kecamatan dan kota (mis. Kemang, Jakarta Selatan)
2	Luas Bangunan (m ²)	Integer	Luas bangunan dalam meter persegi
3	Luas Tanah (m ²)	Integer	Luas tanah dalam meter persegi
4	Kamar Tidur	Integer	Jumlah kamar tidur (1–6)
5	Kamar Mandi	Integer	Jumlah kamar mandi (1–4)
6	Garasi	Integer/Biner	Ketersediaan garasi (0 = Tidak Ada, 1 = Ada)
7	Jumlah Lantai	Integer	Jumlah lantai bangunan (1–4)
8	Harga (Miliar Rp)	Float	Harga rumah dalam miliar rupiah

Sumber: (Penulis, 2026)

Sebelum dianalisis, seluruh kolom numerik yang masih berformat teks (misalnya “154 m²” atau “Rp 6,38 Miliar”) dikonversi menjadi tipe numerik, sedangkan kolom lokasi diproses melalui label encoding dan one-hot encoding agar dapat digunakan sebagai variabel prediktor dalam regresi linear. Hasil pemeriksaan missing value menunjukkan bahwa seluruh kolom dataset tidak memiliki



nilai yang hilang (0%), sehingga tidak diperlukan proses imputasi atau penghapusan baris akibat data kosong.

Deteksi pencilan dilakukan pada tiga kolom numerik yang paling berpotensi memengaruhi validitas model, yaitu luas bangunan, luas tanah, dan harga, menggunakan metode Interquartile Range (IQR). Hasil perhitungan disajikan pada Tabel 2.

Tabel 2. Deteksi Outlier Menggunakan Metode IQR

Kolom	Q1	Q3	IQR	Batas Bawah	Batas Atas	Jumlah Outlier
Luas Bangunan	88	248	160	-152	488	14
Luas Tanah	96	316	220	-234	646	19
Harga (Miliar Rp)	1,48	5,85	4,37	-5,08	12,41	31

Sumber: (Penulis, 2026)

Baris data yang memiliki nilai di luar batas bawah atau batas atas pada salah satu kolom tersebut dihapus dari dataset. Setelah proses pembersihan, dataset akhir yang digunakan untuk pemodelan regresi berjumlah 692 data valid.

3.2 Statistik Deskriptif

Statistik deskriptif dari 692 observasi valid menunjukkan variasi harga yang cukup signifikan, dengan rata-rata harga rumah sebesar Rp7,40 miliar dan simpangan baku Rp3,766 miliar. Rata-rata luas tanah sebesar 179,00 m² dan luas bangunan 140,18 m² menunjukkan bahwa sebagian besar properti dalam dataset berukuran menengah, dengan rata-rata 3,90 kamar tidur dan 3,38 kamar mandi. Sebanyak 84% properti dalam dataset memiliki garasi (mean = 0,84). Dari sisi lokasi, Jakarta Selatan merupakan wilayah dengan proporsi data terbesar (31%), diikuti Jakarta Utara (21%), Jakarta Timur (12%), dan Jakarta Pusat (10%), sedangkan Jakarta Barat (26%) berperan sebagai kategori baseline (referensi) dalam model regresi.

3.3 Persamaan Model Regresi Linear Berganda

Hasil analisis regresi menggunakan metode Ordinary Least Squares pada SPSS menghasilkan persamaan prediksi harga rumah sebagai berikut:

$$\hat{Y} = 0,157 + 2,296(\text{Lok_Jakarta_Pusat}) + 0,726(\text{Lok_Jakarta_Selatan}) - 1,496(\text{Lok_Jakarta_Timur}) + 0,872(\text{Lok_Jakarta_Utara}) + 0,011(\text{Luas_Bangunan_m2}) + 0,021(\text{Luas_Tanah_m2}) - 0,137(\text{Kamar_Tidur}) + 0,490(\text{Kamar_Mandi}) + 0,351(\text{Garasi}) + 0,044(\text{Jumlah_Lantai})$$

Dengan Jakarta Barat sebagai kategori pembanding, rumah di Jakarta Pusat diprediksi paling mahal (+Rp2,296 miliar), diikuti Jakarta Utara (+Rp0,872 miliar) dan Jakarta Selatan (+Rp0,726 miliar), sementara Jakarta Timur justru lebih murah (-Rp1,496 miliar). Dari sisi karakteristik fisik, setiap penambahan 1 m² luas tanah menaikkan harga sebesar Rp0,021 miliar dan luas bangunan Rp0,011 miliar, sedangkan penambahan satu unit kamar mandi menaikkan harga sebesar Rp0,490 miliar. Sebaliknya, jumlah kamar tidur, garasi, dan jumlah lantai menunjukkan koefisien yang tidak signifikan secara statistik (Sig. > 0,05), sehingga arah pengaruhnya tidak dapat digeneralisasi.

3.4 Uji Kebaikan Model (Model Summary dan Uji F)

Tabel 3. Ringkasan Model (Model Summary)

R	R Square (R ²)	Adjusted R ²	Std. Error of the Estimate
0,742	0,550	0,543	2,545 (miliar Rp)

Sumber: (Penulis, 2026)



Tabel 4. Hasil Uji Simultan (ANOVA / Uji F)

Sumber Variasi	Sum of Squares	df	Mean Square	F	Sig.
Regression	5.388,348	10	538,835	83,185	<0,001
Residual	4.411,233	681	6,478		
Total	9.799,581	691			

Sumber: (Penulis, 2026)

Nilai koefisien determinasi (R^2) sebesar 0,550 dan Adjusted R^2 sebesar 0,543 menunjukkan bahwa model mampu menjelaskan 55% variasi harga rumah di wilayah DKI Jakarta, sedangkan 45% sisanya dipengaruhi oleh faktor lain di luar model, seperti kondisi bangunan, usia properti, aksesibilitas transportasi, dan kondisi makroekonomi. Standard Error of the Estimate sebesar 2,545 miliar menunjukkan rata-rata penyimpangan prediksi model terhadap harga aktual. Uji F menghasilkan nilai F hitung sebesar 83,185 dengan signifikansi $< 0,001$, jauh lebih besar dari F tabel ($\approx 1,87$ pada $df_1 = 10$ dan $df_2 = 681$), sehingga H_0 ditolak dan disimpulkan bahwa seluruh variabel independen secara bersama-sama berpengaruh signifikan terhadap harga rumah.

3.5 Uji Signifikansi Parsial (Uji t)

Hasil uji t pada Tabel 5 menunjukkan bahwa dari sepuluh variabel independen yang diuji, tujuh variabel berpengaruh signifikan secara parsial (Sig. $< 0,05$), yaitu Lok_Jakarta_Pusat, Lok_Jakarta_Selatan, Lok_Jakarta_Timur, Lok_Jakarta_Utara, Luas_Bangunan_m2, Luas_Tanah_m2, dan Kamar_Mandi, sedangkan tiga variabel lainnya (Kamar_Tidur, Garasi, dan Jumlah_Lantai) tidak berpengaruh signifikan. Variabel luas tanah memiliki koefisien beta terstandarisasi terbesar ($\beta = 0,496$), menjadikannya prediktor paling dominan terhadap harga rumah di wilayah DKI Jakarta.

Tabel 5. Hasil Uji Signifikansi Parsial (Uji t)

Variabel	Koefisien (B)	Sig.	Keterangan
(Constant)	0,157	$>0,05$	Tidak Signifikan
Lok_Jakarta_Pusat	2,296	$<0,001$	Signifikan
Lok_Jakarta_Selatan	0,726	0,006	Signifikan
Lok_Jakarta_Timur	-1,496	$<0,001$	Signifikan
Lok_Jakarta_Utara	0,872	0,002	Signifikan
Luas_Bangunan_m2	0,011	$<0,001$	Signifikan
Luas_Tanah_m2	0,021	$<0,001$	Signifikan (paling dominan)
Kamar_Tidur	-0,137	0,536	Tidak Signifikan
Kamar_Mandi	0,490	0,012	Signifikan
Garasi	0,351	0,191	Tidak Signifikan
Jumlah_Lantai	0,044	0,748	Tidak Signifikan

Sumber: (Penulis, 2026)

3.6 Uji Asumsi Klasik

Pengujian asumsi klasik menunjukkan bahwa model regresi telah memenuhi syarat sebagai Best Linear Unbiased Estimator (BLUE). Histogram residual terstandarisasi membentuk kurva mendekati lonceng dengan mean = $-1,26E-15$ dan standar deviasi = 0,993, yang dikonfirmasi oleh Normal P-P Plot yang mengikuti garis diagonal tanpa deviasi substansial, sehingga asumsi normalitas terpenuhi. Diagram sebar residual tidak menunjukkan pola sistematis (seperti bentuk



corong), sehingga asumsi homoskedastisitas juga terpenuhi. Pengujian multikolinearitas menunjukkan bahwa empat variabel memiliki nilai VIF antara 5 dan 10, yang mengindikasikan multikolinearitas moderat namun belum mengancam validitas model secara serius, karena tidak ditemukan nilai VIF di atas 10. Terakhir, uji autokorelasi menggunakan statistik Durbin-Watson menghasilkan nilai 2,017, sangat mendekati 2, sehingga tidak ditemukan indikasi autokorelasi pada residual model.

Statistik residual menunjukkan nilai prediksi (predicted value) berkisar antara Rp0,99 miliar hingga Rp16,03 miliar dengan rata-rata Rp7,40 miliar, sama dengan rata-rata harga aktual, yang mengonfirmasi model tidak mengandung bias sistematis. Residual aktual berkisar antara -Rp6,605 miliar hingga +Rp8,991 miliar. Nilai Cook's Distance maksimum sebesar 0,020 dan leverage terpusat maksimum sebesar 0,040 (rata-rata 0,014) menunjukkan bahwa tidak ada satu pun observasi yang mendominasi atau mengubah parameter model secara signifikan.

3.7 Implementasi Sistem Prediksi Berbasis Web

Persamaan regresi yang dihasilkan selanjutnya diimplementasikan ke dalam aplikasi berbasis web menggunakan bahasa pemrograman PHP untuk logika aplikasi dan antarmuka pengguna, serta basis data MySQL untuk menyimpan data properti dan riwayat prediksi. Sistem dirancang menggunakan pemodelan Unified Modeling Language (UML) dengan dua aktor utama, yaitu Admin dan Pengguna. Selain fitur prediksi harga rumah, sistem ini dilengkapi dengan fitur rekomendasi properti berdasarkan anggaran, simulasi Kredit Pemilikan Rumah (KPR), komparasi harga antarwilayah, riwayat prediksi, serta dashboard admin untuk mengelola koefisien model secara dinamis. Dengan demikian, pengguna, baik masyarakat umum, calon pembeli, investor, maupun agen properti, dapat memperoleh estimasi harga rumah secara interaktif dan real-time berdasarkan spesifikasi properti yang dimasukkan.

3.8 Kelebihan dan Keterbatasan Penelitian

Penelitian ini memiliki sejumlah kelebihan, di antaranya integrasi menyeluruh mulai dari pengumpulan data otomatis, analisis statistik, pemodelan regresi, hingga implementasi aplikasi web; cakupan wilayah yang lebih luas dan representatif karena mencakup seluruh lima wilayah administratif DKI Jakarta; model yang telah lolos seluruh uji asumsi klasik OLS; serta kualitas data yang baik tanpa missing value. Meskipun demikian, penelitian ini juga memiliki sejumlah keterbatasan, yaitu daya jelas model yang tergolong sedang ($R^2 = 0,550$) sehingga 45% variasi harga belum dapat dijelaskan oleh model; margin kesalahan prediksi yang masih cukup besar (Standard Error of Estimate = 2,545 miliar); indikasi multikolinearitas moderat pada empat variabel; tiga variabel yang tidak berkontribusi signifikan terhadap model; keterbatasan pendekatan regresi linear dalam menangkap hubungan non-linear antarvariabel; evaluasi model yang belum menggunakan skema data latih dan data uji terpisah; serta sumber data yang hanya berasal dari satu platform (Rumah123.com) dan bersifat cross-sectional sehingga belum mampu menangkap tren harga dari waktu ke waktu.

4. KESIMPULAN

Berdasarkan hasil analisis dan pembahasan, dapat ditarik beberapa kesimpulan sebagai berikut.

- Data properti residensial di lima wilayah administratif DKI Jakarta berhasil dikumpulkan melalui teknik web scraping menggunakan Python (BeautifulSoup dan Requests) dari platform Rumah123.com, menghasilkan 750 data awal dengan delapan atribut yang seluruhnya bebas dari missing value.
- Melalui metode regresi linear berganda dengan estimasi Ordinary Least Squares pada 692 data valid (setelah penanganan outlier menggunakan metode IQR), dihasilkan persamaan model prediksi harga rumah yang terbukti valid secara statistik berdasarkan uji F ($F = 83,185$; $\text{Sig.} < 0,001$) dan telah memenuhi seluruh asumsi klasik OLS, yaitu normalitas, homoskedastisitas, non-multikolinearitas serius, dan non-autokorelasi (Durbin-Watson = 2,017).
- Harga rumah di wilayah DKI Jakarta secara signifikan dipengaruhi oleh tujuh dari sepuluh variabel yang diuji, yaitu lokasi wilayah, luas bangunan, luas tanah, dan jumlah kamar mandi, dengan luas tanah sebagai prediktor paling dominan, sedangkan jumlah kamar tidur, ketersediaan garasi, dan jumlah lantai tidak terbukti berpengaruh signifikan secara parsial.



JRIIN : Jurnal Riset Informatika dan Inovasi
Volume 4, No. 7 Tahun 2026
ISSN 3025-0919 (media online)
Hal 1905-1912

- d. Tingkat akurasi model regresi linear berganda yang dihasilkan tergolong sedang, dengan koefisien determinasi R^2 sebesar 0,550, yang berarti model mampu menjelaskan 55% variasi harga rumah, sementara 45% sisanya dipengaruhi oleh faktor di luar model seperti kondisi bangunan, usia properti, aksesibilitas, dan kondisi makroekonomi.
- e. Persamaan regresi yang dihasilkan berhasil diimplementasikan ke dalam aplikasi berbasis web menggunakan PHP dan MySQL, dilengkapi fitur prediksi harga, rekomendasi properti berdasarkan anggaran, simulasi KPR, komparasi harga antarwilayah, riwayat prediksi, serta dashboard admin, sehingga dapat digunakan secara interaktif oleh masyarakat umum, investor, dan pelaku industri properti.
- f. Penelitian ini berhasil menjembatani kesenjangan penelitian sebelumnya yang umumnya terbatas pada cakupan wilayah sempit dan belum mengintegrasikan pengumpulan data otomatis, pemodelan statistik, serta antarmuka digital yang dapat diakses langsung oleh masyarakat.

Mengingat R^2 penelitian ini baru mencapai 55%, penelitian selanjutnya disarankan untuk memperluas sumber data ke platform properti lain, menambahkan variabel eksternal seperti suku bunga KPR dan jarak fasilitas umum, menerapkan skema data latih-uji atau validasi silang, serta mempertimbangkan metode pembelajaran mesin lain seperti Random Forest atau Gradient Boosting untuk menangkap hubungan non-linear antarvariabel secara lebih baik.

UCAPAN TERIMA KASIH

Penulis mengucapkan puji syukur kepada Tuhan Yang Maha Esa atas rahmat dan karunia-Nya sehingga penelitian ini dapat diselesaikan dengan baik. Ucapan terima kasih disampaikan secara khusus kepada Ibu Siti Khotijah, S.Kom., M.M.S.I. selaku pembimbing materi dan Ibu Dr. Siwi Puji Astuti, M.Pd. selaku pembimbing teknik atas arahan, kritik, dan saran yang membangun selama penyusunan penelitian ini. Penulis juga mengucapkan terima kasih kepada Program Studi Teknik Informatika, Fakultas Teknik dan Ilmu Komputer, Universitas Indraprasta PGRI, serta kepada Rumah123.com yang data listingnya digunakan sebagai sumber data penelitian ini.

REFERENCES

- Athallah, M. R., & Rozi, A. F. (2023). Implementasi Data Mining Untuk Prediksi Peramalan Penjualan Produk Hj Karpel Menggunakan Metode Linear Regression. *Jurnal Sains Dan Teknologi (JSIT)*, 2(3), 180–187. <https://doi.org/10.47233/jsit.v2i3.550>
- Chen, Q., & Qi, J. (2023). How much should we trust R^2 and adjusted R^2 : evidence from regressions in top economics journals and Monte Carlo simulations. *Journal of Applied Economics*, 26(1). <https://doi.org/10.1080/15140326.2023.2207326>
- Ghazali, I. (2018). Aplikasi Analisis Multivariate dengan Program IBM SPSS 25 (Edisi ke-9). Badan Penerbit Universitas Diponegoro.
- Hakim, H., Kamil, D., & Alatas, B. (2025). Pendekatan Machine Learning untuk Estimasi Harga Rumah dengan Regresi Linier. *ALPHA: Journal of Science and Technology*, 1(1), 18–22. <https://doi.org/10.70716/alpha.v1i1.99>
- Hallan, R. R., & Fajri, I. N. (2025). Prediksi Harga Rumah menggunakan Machine Learning Algoritma Regresi Linier. *Jurnal Teknologi Dan Sistem Informasi Bisnis*, 7(1), 57–62. <https://doi.org/10.47233/jteksis.v7i1.1732>
- Husdi, H., & Dalai, H. (2023). Penerapan Metode Regresi Linear Untuk Prediksi Jumlah Bahan Baku Produksi Selai Bilfagi. *Jurnal Informatika*, 10(2), 129–135. <https://doi.org/10.31294/inf.v10i2.14129>
- Junaidi, J., Cholisoh, N., & Nurjaman, D. R. (2024). Pembelajaran Mesin Untuk Prediksi Harga Rumah Dengan Metode Regresi Linier Berganda. *Journal Sensi*, 10(1), 79–89. <https://doi.org/10.33050/sensi.v10i1.3119>
- Kutner, M. H., Nachtsheim, C. J., & Neter, J. (2014). *Applied Linear Regression Models (International Revised Edition)*. McGraw-Hill.
- Mardiatmoko, G. (2020). The Importance of the Classical Assumption Test in Multiple Linear Regression Analysis (A Case Study of the Preparation of the Allometric Equation of Young Walnuts). *Barekeng*, 14(3), 333–342. <https://doi.org/10.30598/barekengvol14iss3pp333-342>
- Maula, I., Hasanah, L. U., & Tholib, A. (2023). Analisis Prediksi Harga Rumah Di Jabodetabek Menggunakan Multiple Linear Regression. *Jurnal Informatika Kaputama (JIK)*, 7(2), 216–224. <https://doi.org/10.59697/jik.v7i2.135>



JRIIN : Jurnal Riset Informatika dan Inovasi
Volume 4, No. 7 Tahun 2026
ISSN 3025-0919 (media online)
Hal 1905-1912

- Maulud, D., & Abdulazeez, A. M. (2020). A Review on Linear Regression Comprehensive in Machine Learning. *Journal of Applied Science and Technology Trends*, 1(2), 140–147. <https://doi.org/10.38094/jastt1457>
- Miftahuljannah, Sunge, A. S., & Zy, A. T. (2023). Analisis Prediksi Penjualan Dengan Metode Regresi Linear Di PT. Eagle Industry Indonesia. *Jurnal Informatika Teknologi Dan Sains (Jinteks)*, 5(3), 398–403. <https://doi.org/10.51401/jinteks.v5i3.3325>
- Mu'tashim, M. L., Muhayat, T., Damayanti, S. A., Zaki, H. N., & Wirawan, R. (2021). Analisis Prediksi Harga Rumah Sesuai Spesifikasi Menggunakan Multiple Linear Regression. *Informatik: Jurnal Ilmu Komputer*, 17(3), 238. <https://doi.org/10.52958/iftk.v17i3.3635>
- Niamilah, A., Alfin, A. A., & Kurniasari, I. (2023). Siklus Hidup Pengembangan Sistem Basis Data Pada Sistem Informasi Buku Tamu di Badan Pusat Statistik Kabupaten Kediri Menggunakan MySQL. *Jurnal Nasional Komputasi Dan Teknologi Informasi (JNKTI)*, 6(1), 115–121. <https://doi.org/10.32672/jnkti.v6i1.5830>
- Novianty, D., Palasara, N. D., & Qomaruddin, M. (2021). Algoritma Regresi Linear pada Prediksi Permohonan Paten yang Terdaftar di Indonesia. *Jurnal Sistem Dan Teknologi Informasi (Justin)*, 9(2), 81. <https://doi.org/10.26418/justin.v9i2.43664>
- Nuris, N. (2024). Analisis Prediksi Harga Rumah Pada Machine Learning Metode Regresi Linear. *Explore*, 14(2), 108–112. <https://doi.org/10.35200/ex.v14i2.123>
- Rizquina, A. Z., & Ratnasari, C. I. (2023). Implementasi Web Scraping untuk Pengambilan Data Pada Website E-Commerce. *Jurnal Teknologi Dan Sistem Informasi Bisnis*, 5(4), 377–383. <https://doi.org/10.47233/jteksis.v5i4.913>
- Sanjaya, K. D. (2024). Prediksi Harga Rumah dengan Metode Regresi Linear dan Support Vector Regression di Daerah Tebet Jakarta Selatan. *Versi Cetak*, 19(2).
- Sholeh, M., Nurnawati, E. K., & Lestari, U. (2023). Penerapan Data Mining dengan Metode Regresi Linear untuk Memprediksi Data Nilai Hasil Ujian Menggunakan RapidMiner. *JISKA (Jurnal Informatika Sunan Kalijaga)*, 8(1), 10–21. <https://doi.org/10.14421/jiska.2023.8.1.10-21>
- Yasin, V. (2021). Tools Rekayasa Perangkat Lunak dalam Membuat Pemodelan Desain Menggunakan Unified Modeling Language (UML). *Tridharmadimas: Jurnal Pengabdian Kepada Masyarakat Jayakarta*, 1(2), 139. <https://doi.org/10.52362/tridharmadimas.v1i2.666>