

Analisis Faktor Penentu Kategori Harga Rumah di Kota Tangerang Selatan Menggunakan *Web Crawling* dan Regresi Logistik Multinomial

¹Muhammad Iyad Irviansyah, ²Vanesya Claresta, ³Marshanda Anabella, ⁴Muhammad Rifqo Saputra, ⁵Rido Dwi Kurniawan, ⁶Muh. Masri Sari

^{1,2,3,4,5,6}Program Studi Sistem Informasi, Universitas Pradita, Tangerang

E-mail: ¹muhammad.iyad@student.pradita.ac.id, ²vanesya.claresta@student.pradita.ac.id,
³marshanda.anabella@student.pradita.ac.id, ⁴muhammad.rifqo@student.pradita.ac.id,
⁵rido.dwi@pradita.ac.id, ⁶muh.masri@pradita.ac.id,

ABSTRAK

Penelitian ini bertujuan mengidentifikasi faktor-faktor yang memengaruhi kategori harga rumah di Kota Tangerang Selatan menggunakan data digital berbasis *web crawling* dari situs properti daring. Permasalahan utama penelitian adalah bagaimana karakteristik fisik, fasilitas, dan lokasi memengaruhi probabilitas rumah berada pada kategori harga tertentu. Data dikumpulkan secara otomatis menggunakan teknik *web crawling* dan setelah proses pembersihan serta validasi diperoleh 1.264 data rumah yang layak dianalisis. Harga rumah diklasifikasikan ke dalam empat kategori, yaitu Ekonomis, Standar, Mewah, dan Eksklusif, menggunakan pendekatan kuartil. Analisis dilakukan dengan Regresi Logistik Multinomial untuk memodelkan peluang relatif setiap kategori harga berdasarkan variabel luas tanah, luas bangunan, jumlah kamar tidur, jumlah kamar mandi, keberadaan garasi, dan lokasi kecamatan. Hasil penelitian menunjukkan bahwa luas tanah, luas bangunan, jumlah kamar mandi, dan keberadaan garasi berpengaruh signifikan terhadap kategori harga rumah, sedangkan jumlah kamar tidur dan lokasi kecamatan tidak menunjukkan pengaruh signifikan setelah dikontrol oleh variabel fisik lainnya. Model yang dibangun signifikan secara simultan dan memiliki tingkat akurasi klasifikasi sebesar 64,8%. Kontribusi utama penelitian ini terletak pada integrasi *web crawling* dan Regresi Logistik Multinomial sebagai pendekatan klasifikasi harga rumah berbasis data digital, yang memberikan kerangka empiris bagi analisis pasar perumahan dan pengembangan sistem valuasi properti otomatis.

Kata kunci : *Web Crawling*, Regresi Logistik Multinomial, Kategori Harga Rumah, Data Digital, Pasar Perumahan

ABSTRACT

This study aims to identify factors influencing housing price categories in South Tangerang City using digital data obtained through web crawling from online property platforms. The research addresses how physical attributes, facilities, and location affect the probability of a house belonging to a specific price category. Data were automatically collected via web crawling, and after data cleaning and validation, 1,264 housing records were retained for analysis. Housing prices were classified into four categories—Economical, Standard, Luxury, and Exclusive—using a quartile-based approach. Multinomial Logistic Regression (MLR) was applied to model relative

probabilities based on land area, building area, number of bedrooms, number of bathrooms, garage availability, and district location. The results indicate that land area, building area, number of bathrooms, and garage availability significantly influence housing price categories, while the number of bedrooms and district location are not significant after controlling for physical characteristics. The model is statistically significant and achieves a classification accuracy of 64.8%. The main contribution of this study lies in the integration of web crawling and Multinomial Logistic Regression for housing price classification, offering a data-driven framework to support housing market analysis and automated property valuation systems.

Keyword : *Web Crawling, Multinomial Logistic Regression, Housing Price Categories, Digital Data, Housing Market*

1. PENDAHULUAN

Kota Tangerang Selatan memiliki konteks sosial-ekonomi yang kuat, tercermin dari Indeks Pembangunan Manusia (IPM) 82,28 pada 2023 (BPS Kab. Tangerang, 2024), dan pada saat yang sama dinamika pasar tinggi terlihat dari kenaikan permintaan rumah sekunder sebesar 84% pada kuartal terakhir 2023 (Pinhome, 2024), sehingga mengindikasikan adanya stratifikasi segmen harga yang perlu dipetakan secara terukur. Dinamika juga tampak pada sektor sewa karena inventori rumah sewa nasional meningkat 104% (YoY) sepanjang 2023 dan Tangerang Selatan termasuk lokasi teratas untuk menyewa rumah (Pinhome, 2024), sehingga analisis faktor penentu kategori harga penting untuk menjelaskan perbedaan karakteristik rumah antar segmen secara kuantitatif.

Penelitian terdahulu umumnya menempatkan penentu harga rumah pada atribut fisik, lingkungan permukiman, dan lokasi, dengan pengaruh yang dapat berubah mengikuti dinamika spasial dan pasar (Liu et al., 2025), namun sebagian studi membutuhkan indikator lokasi rinci berbasis jarak dan data geospasial sehingga kurang sejalan dengan keterbatasan variabel pada data listing publik (Fikri et al., 2024), sementara riset berbasis listing menunjukkan dominasi variabel fisik dan fasilitas dasar serta

potensi anomali ketika faktor eksternal tidak tercakup (Aridho et al., 2025).

Berdasarkan uraian tersebut, penelitian ini bertujuan mengukur faktor pembeda kategori harga rumah di Kota Tangerang Selatan pada data listing dari website properti daring dengan regresi logistik multinomial, sehingga peluang masuk kategori Ekonomis, Standar, Mewah, atau Eksklusif dapat dijelaskan melalui luas tanah, luas bangunan, kamar tidur, kamar mandi, garasi, dan kecamatan, serta hasilnya bermanfaat bagi pemerintah sebagai dasar kebijakan keterjangkauan berbasis segmentasi wilayah, bagi pelaku bisnis untuk strategi harga dan produk yang tepat sasaran, dan bagi platform properti daring untuk meningkatkan rekomendasi harga, pengelompokan listing, dan deteksi harga yang tidak wajar.

Rumusan masalah dalam penelitian ini adalah sebagai berikut:

- 1) Bagaimana karakteristik dan distribusi harga rumah di Tangerang Selatan berdasarkan data digital dari situs properti daring?
- 2) Faktor-faktor apa saja yang berpengaruh signifikan terhadap kategori harga rumah (Ekonomis, Standar, Mewah, Eksklusif)?
- 3) Bagaimana model regresi logistik multinomial dapat digunakan untuk memprediksi kategori harga rumah berdasarkan variabel-variabel penting yang teridentifikasi?

2. TINJAUAN PUSTAKA

2.1. Faktor Penentu Harga Rumah

Faktor penentu harga rumah adalah elemen pembentuk nilai jual properti, sejalan dengan Kotler dan Keller (2016) bahwa harga sebagai komponen kunci marketing mix merepresentasikan nilai ekonomi, persepsi, daya beli, dan dinamika pasar, serta dipengaruhi atribut fisik (Aridho et al., 2025) dan fasilitas serta lokasi yang membentuk persepsi nilai dan daya saing (Fikri et al., 2024).

2.2 Web Crawling

Web crawling adalah penelusuran dan pengambilan halaman web secara otomatis melalui HTTP atau browser untuk mengumpulkan sumber data yang dapat diekstraksi lewat *web scraping* dan disimpan ke file atau basis data sesuai tujuan (Darmawan et al., 2022).

2.3 Regresi Logistik Multinomial

Regresi Logistik Multinomial memodelkan hubungan prediktor dengan variabel dependen nominal berkategori lebih dari dua dengan membandingkan tiap kategori terhadap kategori referensi, sehingga cocok untuk menganalisis probabilitas rumah masuk kategori harga tertentu berdasarkan variabel independen (Newsom, 2025).

2.4 Penelitian Terdahulu

Penelitian terdahulu konsisten menunjukkan faktor struktural dan lokasi memengaruhi harga rumah, meliputi peran lokasi (Rahadi et al., 2022), kedekatan dengan pusat kota (Fikri et al., 2024), dominasi luas bangunan, luas tanah, dan jumlah kamar (Aridho et al., 2025), serta bukti Regresi Logistik Multinomial bahwa karakteristik fisik dan lokasi menentukan kategori harga (Teoh et al., 2022), sementara penelitian ini memakai data *web crawling* dan Regresi Logistik Multinomial untuk menganalisis

penentu kategori harga rumah di Tangerang Selatan.

2.5 Kerangka Pemikiran



Gambar 2. Algoritma Penulisan Makalah TRANSIT

2.6 Hipotesis Penelitian

Hipotesis penelitian ini menyatakan bahwa faktor fisik, fasilitas, dan lokasi memengaruhi harga rumah di Tangerang Selatan.

- H1: Luas tanah berpengaruh dan signifikan terhadap kategori harga rumah di Tangerang Selatan.
- H2: Luas bangunan berpengaruh dan signifikan terhadap kategori harga rumah di Tangerang Selatan
- H3: Jumlah kamar tidur berpengaruh dan signifikan terhadap kategori harga rumah di Tangerang Selatan
- H4: Jumlah kamar mandi berpengaruh dan signifikan terhadap kategori harga rumah di Tangerang Selatan
- H5: Keberadaan garasi berpengaruh dan signifikan terhadap kategori harga rumah di Tangerang Selatan
- H6: Lokasi berpengaruh signifikan dan terhadap kategori harga rumah di Tangerang Selatan.
- H7: Faktor fisik, fasilitas, dan lokasi secara simultan berpengaruh signifikan terhadap kategori harga rumah di Kota Tangerang Selatan.

2.7 Model Konseptual Penelitian



Gambar 2. Model Konseptual Penelitian

3. METODOLOGI PENELITIAN

3.1 Desain Penelitian

Penelitian kuantitatif ini memakai data *web crawling* dan Regresi Logistik Multinomial untuk mengklasifikasikan kategori harga rumah berdasarkan faktor fisik dan lokasi (Newsom, 2025).

3.2. Sumber Data dan Kualitas Data

Data dikumpulkan daring dari situs properti aktif yang merepresentasikan pasar Kota Tangerang Selatan.

3.3 Metode Pengumpulan Data (*Web Crawling*)

Data dikumpulkan melalui *web crawling* dengan Python (BeautifulSoup, Selenium) lalu disimpan sebagai CSV berisi harga, luas tanah/bangunan, kamar, garasi, dan lokasi dari sumber publik tanpa data pribadi sesuai etika (Oprea & Bâra, 2022).

3.4 Teknik Pengumpulan Data

Data diambil lewat *crawling* Python (Selenium–BeautifulSoup, Chrome Headless) pada Okt–Des 2025, mengekstraksi atribut via CSS/XPath dan *auto-scroll*, menerapkan jeda acak, rotasi *User-Agent*, *backoff* 429, cek *robots.txt*, lalu mengumpulkan judul, URL, harga, luas, kamar, garasi, dan kecamatan (kode ordinal).

3.5 Teknik Analisis Data

Data yang diperoleh akan dianalisis melalui beberapa tahapan berikut:

- 1) Validasi: hapus duplikasi (URL), saring harga tak realistis (<200 juta atau >10 miliar), dan cek manual listing aktif serta sesuai wilayah.

- 2) Analisis Deskriptif: Menggambarkan karakteristik dan distribusi variabel.
- 3) Variabel dependen (harga) : dikategorikan menggunakan metode kuartil (Q1–Q3) jadi 4 kelas karena robust outlier dan lebih proporsional dibanding k-means yang sensitif nilai ekstrem.
- 4) Variabel lokasi kecamatan: dikodekan ordinal untuk cegah overfitting koordinat dan wakili wilayah administratif.
- 5) Analisis Regresi Logistik Multinomial: uji pengaruh variabel independen dan interpretasi odds kategori.
- 6) Validasi IIA: kategori distinct dan uji Hausman menunjukkan perubahan jumlah kategori tak mengubah proporsi signifikan, sehingga MNL layak.

$$P(Y = k) = \frac{e^{\beta_{0k} + \beta_{1k}X_1 + \beta_{2k}X_2 + \dots + \beta_{pk}X_p}}{1 + \sum_{j=1}^{K-1} e^{\beta_{0j} + \beta_{1j}X_1 + \beta_{2j}X_2 + \dots + \beta_{pj}X_p}}$$

Gambar 4. Rumus Umum Regresi Logistik Multinomial

3.6 Evaluasi Model

Evaluasi model memakai *confusion matrix* untuk menilai akurasi, error rate, dan pola kesalahan antar kategori, serta mempertimbangkan class imbalance agar kualitas prediksi dinilai lebih objektif (Mohammed et al., 2025).

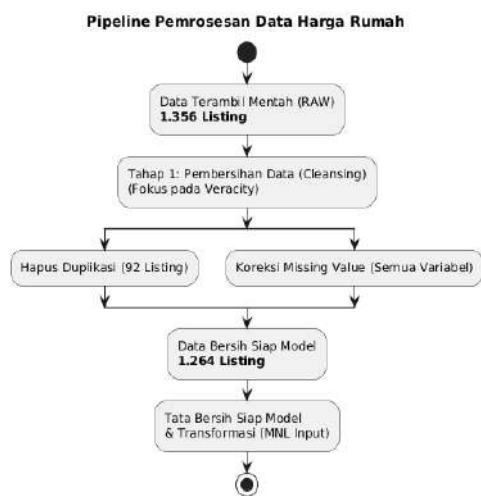
4. HASIL DAN PEMBAHASAN

4.1. Deskripsi Data

4.1.1 Analisis Data Awal

Dataset awal terdiri dari 1.500 listing hasil *web crawling*. Setelah proses data cleansing yang meliputi penghapusan duplikasi, koreksi missing value, dan eliminasi outlier, diperoleh 1.264 data valid yang digunakan dalam analisis Regresi Logistik Multinomial.

4.1.2 Visualisasi Data Pipeline



Gambar 5. Pipeline pemrosesan data.

Gambar 5 menunjukkan alur pemrosesan data mulai dari hasil *web crawling* hingga data final yang siap dianalisis. Pipeline ini menegaskan fokus penelitian pada aspek *veracity* data melalui tahapan pembersihan, validasi, dan transformasi untuk memastikan kualitas dan konsistensi dataset.

4.1.3 Analisis Success rate Crawling Data

Matriks	Jumlah Data	Persentase
Target Crawling	1.500	100%
Success Rate	1.356	90,4%
Error Rate	144	9,6%
Data Duplikat/Null	92	6,78%
Data Valid (Final)	1.264	84,27%

Tabel 1. Success rate crawling data.

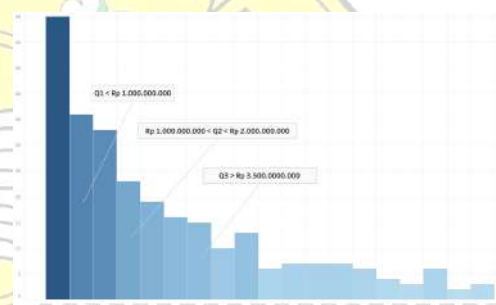
Tingkat keberhasilan *crawling* ditunjukkan pada Tabel 1. Dari target 1.500 listing, terkumpul 1.356 data (90,4%), dengan 92 duplikasi dihapus sehingga menyisakan 1.264 data valid (84,27%). Angka ini menunjukkan bahwa prosedur *crawling* bekerja efektif dan menghasilkan data cukup representatif.

4.1.4 Robustness Model dan Kualitas Data

Data hasil *web crawling* cenderung bersifat *noisy* dan redundan. Melalui proses cleansing dan validasi (Gambar 5), *noise* berhasil diminimalkan sehingga model regresi dibangun dari data yang lebih stabil dan cukup robust untuk menghasilkan estimasi koefisien yang andal.

4.1.5 Analisis Variabel Dependen

Kategori harga rumah dibentuk dengan metode kuartil, dengan Q1 = Rp1 miliar dan Q3 = Rp3,5 miliar. Distribusi data didominasi kategori Standar, sementara kategori lain relatif seimbang, menunjukkan ketidakseimbangan data yang mencerminkan kondisi pasar aktual.



Gambar 6. Distribusi Histogram Harga Rumah.

Kategori Harga Rumah	Rentang Harga (Rp)	Frekuensi
Ekonomis	<1 miliar	195
Standar	1 miliar s/d 2 miliar	479
Mewah	2 miliar s/d 3 miliar	277
Eksklusif	>3,5 miliar	313
Total		1264

Tabel 2. Distribusi Kategori Harga Rumah Berdasarkan Rentang Harga

Tabel 2 menunjukkan kategori Standar (Rp1–Rp2 miliar) paling dominan 479 unit (37,76%), sedangkan Ekonomis (> Rp1 miliar) paling rendah 195 unit (15,34%). Kategori Mewah (Rp2–Rp3,5 miliar) dan Eksklusif (> Rp3,5 miliar)

relatif seimbang. Meski imbalanced, Regresi Logistik Multinomial tetap layak karena kategori terkecil masih memadai (N = 195).

4.1.6 Struktur Variabel Independen

Setelah variabel dependen diklasifikasikan, variabel independen

yang memengaruhi kategori harga dideskripsikan dan dikelompokkan menjadi Fisik, Fasilitas, dan Lokasi, dengan sebagian variabel dikodekan agar dapat dianalisis menggunakan Regresi Logistik Multinomial.

Variabel	Model	Keterangan	Jenis Nilai
Luas Tanah	X ₁	Luas Tanah (m ²)	Nominal
Luas Bangunan	X ₂	Luas Bangunan (m ²)	Nominal
Jumlah Kamar Tidur	X ₃	Jumlah Kamar Tidur	Nominal
Jumlah Kamar Mandi	X ₄	Jumlah Kamar Mandi	Nominal
Garasi	X ₅	Ketersediaan Garasi (1 = Ada, 0 = Tidak Ada)	Nominal
Lokasi	X ₆	Kode Kecamatan (1–4: Ciputat, Pamulang, Pondok Aren, Serpong)	Ordinal

Tabel 3. Variabel Penelitian dan Keterangan Model

Berdasarkan struktur Tabel 3, faktor fisik menjadi penentu utama nilai ekonomi properti karena terkait kapasitas, kenyamanan, dan fungsi ruang, yang direpresentasikan oleh Luas Tanah (X₁) dan Luas Bangunan (X₂). Setelah data cleansing untuk menghapus nilai ekstrem tidak realistis (misalnya < 30 m² atau > 1000 m²), diperoleh statistik deskriptif sebagaimana pada tabel berikut.

Variabel	Min (m ²)	Max (m ²)	Median (m ²)
LT	36	945	112
LB	25	780	120

Tabel 4. Statistik Deskriptif Variabel Fisik (LT dan LB)

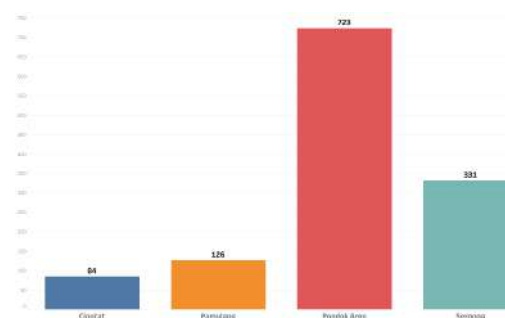
Pada Tabel 4, Luas Tanah (LT) memiliki rentang 36–945 m², mencerminkan variasi tipe rumah dari kecil hingga besar, dengan median 112 m² sebagai ukuran tengah yang dominan di pasar. Luas Bangunan (LB) juga bervariasi 25–780 m², dengan median 120 m².

Variabel	Min	Max	Rata-rata
Jumlah Kamar	1	8	3

Tidur (JKT)			
Jumlah Kamar	1	7	3
Mandi (JKM)			

Tabel 5. Statistik Deskriptif Variabel Fasilitas (JKT dan JKM)

Berdasarkan Tabel 5, kamar tidur 1–8 (rata-rata 3) dan kamar mandi 1–7 (rata-rata 3), menunjukkan pola dominan 3 kamar tidur dan 3 kamar mandi pada rumah tapak menengah perkotaan. Variabel garasi (X₅) dikodekan 1 = ada, 0 = tidak ada dan meski tidak ditampilkan rinci karena dummy, tetap dimasukkan ke model untuk mengukur pengaruhnya terhadap kategori harga.



Gambar 7. Distribusi Jumlah Rumah Berdasarkan Kecamatan di Tangerang Selatan

Variabel lokasi (X_6) dikodekan numerik. Gambar 7 menunjukkan sebaran sampel pada 4 kecamatan yang tidak merata: Pondok Aren (3) 723 (56,3%), Serpong (4) 331 (27,0%), Pamulang (2) 126 (10,1%), dan Ciputat (1) 84 (6,6%). Data terkonsentrasi di Pondok Aren dan Serpong, sehingga perbedaan nilai spasial antarwilayah diuji lanjut dengan Regresi Logistik Multinomial.

4.1.7 Uji Multikolinearitas Data

Sebelum dilakukan pemodelan regresi logistik multinomial, dilakukan uji multikolinearitas menggunakan Variance Inflation Factor (VIF) untuk memastikan tidak terdapat hubungan linier kuat antar variabel independen.

Variabel	VIF
LT (Luas Tanah)	2.1
LB (Luas Bangunan)	3
JKT (Jumlah Kamar Tidur)	2.6
JKM (Jumlah Kamar Mandi)	2.8
Keberadaan Garasi	1.1
Kode Kecamatan	1

Tabel 6. Visualisasi Hasil Uji Multikolinearitas (VIF)

Berdasarkan Tabel 6, seluruh variabel independen memiliki $VIF < 5$. VIF tertinggi pada Luas Bangunan 2,98, diikuti Jumlah Kamar Mandi 2,76, Jumlah Kamar Tidur 2,56, Luas Tanah 2,10, Garasi 1,14, dan Kode Kecamatan 1,03. Ini menunjukkan tidak ada korelasi kuat antarvariabel, sehingga tidak terjadi multikolinearitas signifikan dan semua

variabel layak untuk regresi logistik multinomial.

4.2. Analisis Regresi Logistik Multinomial

4.2.1 Model Regresi Untuk Kategori Harga

Tahap berikutnya adalah membangun model prediksi kategori harga rumah di Tangerang Selatan dengan Regresi Logistik Multinomial, karena variabel dependen memiliki empat kategori (Ekonomis, Standar, Mewah, Eksklusif). Model mengestimasi odds tiap kategori berdasarkan karakteristik fisik dan lokasi, dengan lokasi sebagai dummy kecamatan dan Pondok Aren sebagai baseline. Secara matematis, model dinyatakan sebagai berikut:

$$\ln \left(\frac{P(Y = k)}{P(Y = \text{Ekonomis})} \right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n$$

Gambar 8. Model Matematis Regresi Logistik Multinomial untuk Analisis Harga Rumah

- a) Y : Kategori harga rumah
- b) $P(Y=k)$: Probabilitas rumah termasuk kategori harga ke- k
- c) β_i : Koefisien regresi variabel independen
- d) X_i : Variabel independen seperti luas tanah (LT), luas bangunan (LB), jumlah kamar tidur (JKT), jumlah kamar mandi (JKM), garasi (KG), dan lokasi kecamatan (Cip, Pam, Ser)

4.2.2 Hasil Analisis Regresi

Dalam model regresi logistik multinomial ini, kategori Ekonomis ($Y=1$) ditetapkan sebagai kategori acuan (*baseline*) karena memiliki urutan terkecil berdasarkan semua kategori serta mewakili kelompok harga dasar.

Logit	Variabel	B	Wald	P-value
Y = 2 (Standar)	const	-6.1493*	104.658*	0*
	LT	0.0197*	54.471*	0*
	LB	0.0144*	29.512*	0*
	JKT	-0.2489	3.307	0.069

Y = 3 (Mewah)	JKM	0.5324*	13.567*	0*
	Keberadaan Garasi	0.6994*	12.784*	0*
	Lokasi	0.2016	2.8	0.094
	const	-11.5293*	170.636*	0*
	LT	0.0269*	81.889*	0*
	LB	0.0238*	65.801*	0*
	JKT	-0.3785*	4.615*	0.032*
	JKM	1.2649*	44.008*	0*
	Keberadaan Garasi	1.2985*	18.771*	0*
	Lokasi	0.1847	1.197	0.274
Y = 4 (Eksklusif)	const	3.0765*	32.607*	0*
	LT	-0.0118*	8.336*	0.004*
	LB	-0.0196*	19.143*	0*
	JKT	0.1074	0.348	0.555
	JKM	-0.8305*	18.676*	0*

Tabel 7. Analisis Regresi Logistik Multinomial Mempengaruhi Harga Rumah

Berdasarkan hasil analisis regresi logistik multinomial pada Tabel 7 diperoleh interpretasi sebagai berikut:

1. Luas tanah signifikan ($p < 0,05$) pada Standar & Mewah (positif) dan Eksklusif (negatif); pengaruh melemah di segmen tertinggi.
2. Luas bangunan: β positif signifikan ($p < 0,05$) pada Standar & Mewah; β negatif signifikan ($p < 0,05$) pada Eksklusif.
3. Luas bangunan yang lebih besar meningkatkan peluang masuk kategori harga lebih tinggi, tetapi pada kelas Eksklusif tidak lagi membedakan. Luas bangunan tetap faktor fisik paling konsisten menaikkan harga rumah.
4. Kamar tidur negatif; tidak signifikan pada Standar & Eksklusif, tetapi negatif signifikan pada Mewah ($p = 0,032$); efek utama terlihat di menengah-atas dan selain itu terwakili luas bangunan.
5. Kamar mandi signifikan ($p < 0,05$) pada Standar & Mewah (positif) dan Eksklusif (negatif); menambah peluang naik kategori, tetapi di premium faktor lain lebih menentukan.
6. Garasi signifikan ($p < 0,05$) pada Standar & Mewah (positif), tidak signifikan pada Eksklusif; penting di

kelas menengah, kurang membedakan di Eksklusif karena hampir semua ada.

7. Lokasi (kecamatan) tidak signifikan ($p\text{-value} > 0,05$) di semua kategori; perbedaan antar-kecamatan relatif homogen setelah kontrol variabel lain.

4.3 Uji Signifikansi Simultan

Uji signifikansi simultan bertujuan menilai apakah seluruh variabel independen secara bersama-sama memengaruhi kategori harga rumah. Pengujian dilakukan dengan *Likelihood Ratio Test* (G^2) melalui perbandingan model penuh dan null model.

- a. H_0 = Semua koefisien $\beta = 0$ (tidak ada pengaruh variabel independen)
- b. H_1 = Minimal satu koefisien $\beta \neq 0$ (ada pengaruh signifikan)

Statistik Uji	Nilai	p-value	Kesimpulan
G^2 (LR Chi-Square)	1488.37	< 0.001	H_0 ditolak → Model signifikan

Tabel 8. Uji Signifikansi Simultan Seluruh Variabel

Uji Likelihood Ratio pada tabel 8 menunjukkan bahwa seluruh variabel independen secara simultan berpengaruh

signifikan terhadap kategori harga rumah ($p < 0,001$). Artinya, model regresi yang dibangun memiliki kekuatan prediktif secara keseluruhan.

4.4 Uji Model (*Goodness of Fit*)

Uji kecocokan model dilakukan untuk menilai sejauh mana model regresi logistik multinomial mampu menjelaskan variasi kategori harga rumah berdasarkan variabel-variabel independen yang digunakan. Beberapa ukuran kelayakan model (*Goodness of Fit*) ditunjukkan pada Tabel berikut.

Statistik Uji	Nilai	p-value
Likelihood Ratio Chi-Square (G^2)	1365.605	< 0.001
McFadden R^2	0.4048	N/A
Cox & Snell R^2	0.6605	N/A
Nagelkerke R^2	0.7097	N/A
Pearson Chi-Square	3014.161	1

Kategori	Ekonomis	Standar	Mewah	Eksklusif	Total Aktual (N)	Recall (%)
Ekonomis	93	97	5	0	195	47.70%
Standar	42	383	46	8	479	80.00%
Mewah	1	128	101	47	277	36.50%
Eksklusif	0	14	57	242	313	77.30%
Total Prediksi	136	622	209	297	1264	Akurasi Total: 64.8%
Precision (%)	68.40%	61.60%	48.30%	81.50%	-	-

Tabel 10. Tabel *Confusion Matrix* MNL

Model Regresi Logistik Multinomial mencapai akurasi 64,8% pada data pengujian (Tabel 10) dengan performa yang bervariasi antar kategori. Kategori Standar memiliki *Recall* tertinggi (80,0%), sedangkan kategori Eksklusif menunjukkan *Precision* tertinggi (81,5%). Sebaliknya, kategori Mewah memiliki *Recall* terendah (36,5%), menandakan banyak kesalahan klasifikasi ke kategori Standar dan Eksklusif akibat ketidakseimbangan data dan tumpang tindih karakteristik antar kelas.

Akurasi Klasifikasi	64.80%	N/A
---------------------	--------	-----

Tabel 9. Uji Model (*Goodness of Fit*)

Berdasarkan Tabel 9, model signifikan secara keseluruhan dengan $G^2 = 1365,60$ dan $p\text{-value} < 0,001$, sehingga variabel independen berpengaruh simultan terhadap kategori harga rumah. Daya jelaskan model baik dengan McFadden $R^2 = 0,4048$, Cox & Snell $R^2 = 0,6605$, dan Nagelkerke $R^2 = 0,7097$. Kesesuaian model juga baik karena Pearson Chi-Square = 3014,161 dengan $p\text{-value} = 1,000$, dan akurasi klasifikasi 64,8% untuk kategori Ekonomis, Standar, Mewah, Eksklusif, sehingga model layak digunakan untuk analisis faktor fisik dan lokasi.

4.5 Evaluasi Performa Model dan Implikasi Bias Klasifikasi

4.6 Temuan Pembahasan

4.6.1 Dominasi Faktor Fungsionalitas dan Fasilitas

Hasil regresi logistik multinomial menunjukkan jumlah kamar mandi dan garasi paling dominan menaikkan peluang rumah masuk kategori harga lebih tinggi, dengan koefisien positif dan signifikan ($p\text{-value} < 0,05$) di hampir semua kategori. Temuan ini menegaskan fasilitas fungsional dan kenyamanan lebih berpengaruh daripada ukuran fisik semata, sejalan dengan Aridho et al.

(2025), serta mencerminkan preferensi urban pada utilitas dan kemudahan sehingga meningkatkan daya tarik dan nilai jual rumah di Tangerang Selatan.

4.6.2 Kontekstualisasi Variabel Fisik dan Lokasi

Luas Tanah (LT) dan Luas Bangunan (LB) secara umum berpengaruh positif dan signifikan terhadap harga rumah. Namun, pada kategori Eksklusif (di atas kuartil 75%), LT dan LB justru menunjukkan koefisien negatif dan signifikan. Temuan ini mengindikasikan bahwa pada segmen perumahan premium di Tangerang Selatan, harga lebih ditentukan oleh atribut non-fisik klaster seperti desain, privasi, keamanan, dan fasilitas premium, sehingga kualitas dan nilai simbolik lebih dominan dibandingkan luasan fisik rumah.

4.6.3 Kontribusi Metodologis

Penelitian ini berkontribusi metodologis melalui penerapan Regresi Logistik Multinomial (MNLogit) pada kategori harga rumah berbasis big data. Validitasnya dikonfirmasi lewat uji asumsi IIA (Bab III) sehingga koefisien pada Tabel 7 tidak bias dan kategori Ekonomis, Standar, Mewah, Eksklusif bersifat independen. Model mampu mengestimasi peluang relatif $\text{Exp}(\beta)$ dengan akurasi 64,8%, signifikan simultan ($G^2 = 1365,60$; $p < 0,001$), memiliki good fit, dan daya jelaskan tinggi dengan Nagelkerke $R^2 = 0,7097$ (lebih dari 70% variasi kategori).

5. KESIMPULAN

Berdasarkan analisis Regresi Logistik Multinomial pada 1.264 data rumah di Tangerang Selatan, luas tanah, luas bangunan, jumlah kamar mandi, dan keberadaan garasi terbukti berpengaruh signifikan terhadap kategori harga rumah. Sementara itu, jumlah kamar tidur dan lokasi kecamatan tidak signifikan setelah dikontrol oleh variabel fisik/fasilitas lainnya. Faktor fasilitas (terutama kamar

mandi dan garasi) menjadi pendorong paling dominan untuk peluang masuk kategori harga lebih tinggi. Model yang dibangun dinilai layak dan kuat, signifikan secara simultan (G^2 dan Nagelkerke R^2 tinggi), serta berpotensi dijadikan modul awal APVS. Namun, studi ini masih terbatas karena belum memasukkan faktor makroekonomi, spasial detail, preferensi lingkungan, dan fitur non-fisik lainnya.

6. SARAN

1. Menambah variabel non-fisik (mis. suku bunga KPR, LTV, jarak ke pusat kota, sosial-ekonomi),
2. Memakai pendekatan spasial/GIS untuk analisis jarak fasilitas publik.
3. Melakukan analisis *time-series* agar menangkap dinamika harga.
4. Membandingkan dengan metode machine learning (*Random Forest/Gradient Boosting*).
5. Mengembangkan model sebagai modul awal dalam *Automated Property Valuation System* (APVS).
6. Menguji peningkatan performa dengan model seperti *XGBoost/CatBoost*.

DAFTAR PUSTAKA

- Aridho, S. M. R., Aswha, A. H. K., Wahyuni, T. D., & Sari, A. A. (2025). *Analisis Faktor yang Mempengaruhi Harga Rumah Menggunakan Decision Tree*. *Seminar Nasional Teknologi Informasi Dan Bisnis* (SENATIB), 2025.
<https://doi.org/10.47701/flbwa603>
- Badan Pusat Statistik Kabupaten Tangerang. (2024). *Kabupaten Tangerang dalam Angka*. <https://tangerangkab.bps.go.id/id/publication/2024/02/28/94f4bf4039fbb911093c8848/kabupaten-tangerang-dalam-angka-2024.html>
- Brown, M. A., Gruen, A., Maldoff, G., Messing, S., Sanderson, Z., & Zimmer, M. (2024). *Web Scraping for Research: Legal, Ethical, Institutional,*

- and Scientific Considerations. <https://doi.org/10.48550/arXiv.2410.23432Focustolearnmore>
- Darmawan, I., Maulana, M., Gunawan, R., & Widiyasono, N. (2022). *Evaluating Web Scraping Performance Using XPath, CSS Selector, Regular Expression, and HTML DOM With Multiprocessing Technical Applications*. <https://doi.org/10.30630/joiv.6.4.1525>
- Fauzi, A., Maulidah, N., Supriyadi, R., Nalatissifa, H., & Diantika, S. (2025). *Prediksi Harga Properti Di Indonesia Menggunakan Algoritma Random Forest*. *RIGGS: Journal of Artificial Intelligence and Digital Business*, 4(1), 43–49. <https://doi.org/10.31004/riggs.v4i1.367>
- Fikri, A., Pravitasari, A. E., & Indraprahasta, G. S. (2024). *Keragaman Spasial Faktor-faktor Penentu Harga Jual Rumah di Kota Depok*. *TATALOKA*, 26(2), 110–125. <https://doi.org/10.14710/tataloka.26.2.110-125>
- Hausman, J., & Mcfadden, D. (1984). *Specification Tests for the Multinomial Logit Model* (Vol. 52, Issue 5). <https://doi.org/10.2307/1910997>
- Huang, C. A., & Lam, T. (2025). *PRITES: An integrative framework for investigating and assessing web-scraped HTTP-response datasets for research applications*. <http://arxiv.org/abs/2511.13773>
- Laney, D. (2001). *3D Data Management: Controlling Data Volume, Velocity, and Variety*. <https://diegonogare.net/wp-content/uploads/2020/08/3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>
- Liu, T., Wang, J., Liu, L., Peng, Z., & Wu, H. (2025). *What Are the Pivotal Factors Influencing Housing Prices? A Spatiotemporal Dynamic Analysis Across Market Cycles from Upturn to Downturn in Wuhan*. *Land*, 14(2). <https://doi.org/10.3390/land14020356>
- Mohammed, S., Naumann, F., & Harmouch, H. (2025). *Step-by-Step Data Cleaning Recommendations to Improve ML Prediction Accuracy*. *Advances in Database Technology - EDBT*, 28(3), 542–554. <https://doi.org/10.48786/edbt.2025.43>
- Newsom, J. (2025). *Multiple Regression and Multivariate Quantitative Methods*. https://web.pdx.edu/~newsomj/cdaclas/ho_multinomial.pdf
- Nursiyono, J. A., & Dewi, D. M. (2021). *Determinan Harga Tanah di Indonesia Menggunakan Big Data (Studi Kasus: www.lamudi.co.id)*. <https://doi.org/10.53686/jp.v11i2.105>
- Oprea, S. V., & Bâra, A. (2022). *Why Is More Efficient to Combine BeautifulSoup and Selenium in Scraping For Data Under Energy Crisis*. <https://doi.org/10.61801/OUAESS.2022.2.2.19>
- Philip Kotler, & Kevin Lane Keller. (2015). *Marketing management* (15th ed.). Pearson Education.
- Pinhome. (2024). *Indonesia Residential Market*. <https://www.pinhome.id/research/market-report/en/indonesia-residential-market-report-2023-and-outlook-2024>
- Rahadi, R. A., Wiryono, S. K., Nainggolan, Y. A., Afgani, K. F., Yaman, R., Azmi, A. S. M. A., Ismail, F. Z., Saputra, J., Rahmawati, D., & Moulyni, A. (2022). *Determining the factors influencing residential property price: A comparative study between Indonesia and Malaysia*. *Canada. Decision Science Letters*, 11, 485–496. <https://doi.org/10.5267/dsl.2022.6.002>
- Septiaznu, J. (2024). *Rancang Bangun Sistem Prediksi Nilai Jual Properti Berbasis Web Menggunakan Regresi Linear*. *Jurnal Sains Dan Teknologi (JSIT)*, 4(1), 27–45. <https://doi.org/10.47233/jsit.v4i1.1464>

- Somantri, A., Hermina, T., & Kurnaeli. (2019). *Pengaruh Diferensiasi Harga dan Word Of Mouth terhadap Minat Beli Konsumen pada Toko Daih Collection* Garut. <https://doi.org/10.52434/jkm.v13i2.2194>
- Teoh, E. Z., Yau, W.-C., Ong, T. S., & Connie, T. (2022). *Explainable housing price prediction with determinant analysis*. International Journal of Housing Markets and Analysis, 16(5), 1021–1045. <https://doi.org/10.1108/IJHMA-02-2022-0025>
- Train, K. (2002). *Discrete Choice Methods with Simulation*. <https://doi.org/10.1017/CBO9780511805271>

