

ANALISIS REGRESI LINIER BERGANDA PENGARUH UMUR, JARAK MRT, DAN TOKO PADA HARGA RUMAH

Woro Isti Rahayu^{1)*}, Laode Muhammad Aznur Syahfaja²⁾, Muhammad Vito Wristawidya³⁾, Kiki Mustaqim⁴⁾

^{1,2,3,4)}Universitas Logistik dan Bisnis Internasional, Fakultas Logistik, Teknologi dan Bisnis, Program Studi Sains Data

woroiisti@ulbi.ac.id^{1)*}, lmaznursyahfajar@gmail.com²⁾, m.vitoariesta@gmail.com³⁾, kiki@ulbi.ac.id⁴⁾

Received: 20 January 2025

Accepted: 4 February 2024

Published: 6 february 2024



*woroiisti@ulbi.ac.id

Kata Kunci: Harga rumah, Umur rumah, Jarak MRT, Jumlah toko, Regresi linear berganda

DSI: Jurnal Data Science Indonesia is licensed under a Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0).

Abstrak : Penelitian ini menganalisis faktor-faktor yang mempengaruhi harga rumah di Xindian District, New Taipei City, Taiwan, dengan 4.0). menggunakan analisis regresi linear berganda. Variabel independen yang diteliti mencakup umur rumah, jarak ke stasiun MRT, dan jumlah toko terdekat, sementara variabel dependennya adalah harga rumah. Hasil analisis menunjukkan bahwa umur rumah dan jarak ke stasiun MRT memiliki pengaruh negatif terhadap harga rumah, dengan koefisien regresi masing-masing sebesar -0,253 dan -0,005. Sebaliknya, jumlah toko terdekat berpengaruh positif terhadap harga rumah dengan koefisien regresi sebesar 1,297. Model regresi yang diperoleh memiliki nilai R-Square sebesar 0,541, yang menunjukkan bahwa sekitar 54,1% variasi harga rumah dapat dijelaskan oleh variabel independen dalam penelitian ini.

PENDAHULUAN

Rumah adalah bangunan yang berfungsi sebagai tempat tinggal yang layak huni, sarana pembinaan keluarga, cerminan harkat dan martabat penghuninya, dan aset bagi pemiliknya[1]. Di wilayah ibu kota, harga rumah terus meningkat setiap tahun, sejalan dengan meningkatnya harga tanah dan fasilitas umum yang memadai. Peningkatan harga ini dapat dipengaruhi oleh beberapa faktor seperti kondisi fisik rumah, lokasi, serta akses ke fasilitas publik [2], [3], [4], sehingga menuntut adanya sistem prediksi harga rumah yang akurat. Untuk mengatasi ketidakpastian harga rumah, penelitian ini menggunakan metode regresi linear berganda untuk mengidentifikasi faktor-faktor yang berpengaruh signifikan terhadap harga rumah, sehingga dapat membangun model Machine Learning yang mampu memprediksi harga rumah secara lebih tepat [5], [6].

Regresi linear berganda digunakan untuk mengukur pengaruh beberapa variabel independen terhadap harga rumah sebagai variabel dependen [7]. Variabel yang digunakan dalam penelitian ini meliputi umur rumah, jarak ke stasiun MRT, dan jumlah toko terdekat sebagai faktor penentu harga rumah[8], [9]. Berdasarkan penelitian sebelumnya, ditemukan bahwa umur rumah dan jarak ke MRT cenderung memiliki pengaruh negatif terhadap harga rumah, sedangkan jumlah toko terdekat memberikan dampak positif [10]. Dengan membangun model prediksi berbasis regresi ini, penelitian diharapkan dapat menghasilkan alat prediksi yang berguna bagi calon pembeli dalam memperkirakan harga rumah secara lebih terukur.

TINJAUAN LITERATUR

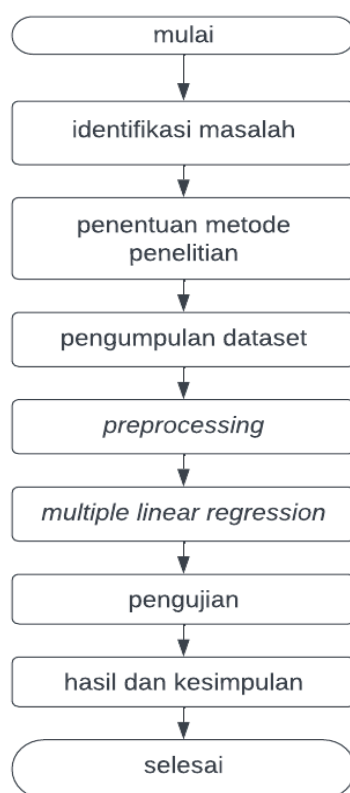
Rahayuningtyas, Evi Febrion, et al. (2021) melakukan penelitian yang berjudul "Prediksi Harga Rumah Menggunakan General Regression Neural Network" dalam Jurnal Informatika Vol. 8 No. 1. Penelitian tersebut menunjukkan bahwa penggunaan General Regression Neural Network (GRNN) mampu memberikan akurasi yang baik dalam memprediksi harga rumah dengan nilai Mean Absolute Percentage Error (MAPE) sebesar 0,045. Hasil penelitian menunjukkan bahwa metode ini cukup akurat, namun kurang optimal dalam menangani data dengan hubungan linier yang kompleks [6]. Mu'tashim et al. menggunakan regresi linear berganda untuk memprediksi harga rumah berdasarkan spesifikasi rumah dan menemukan bahwa model ini mampu memberikan interpretasi yang lebih jelas terhadap hubungan antara variabel-variabel yang memengaruhi

harga rumah [11]. Chen menguji stasioneritas harga rumah di China dan menemukan bahwa terdapat faktor makroekonomi yang memengaruhi fluktuasi harga rumah.

Berbeda dengan penelitian sebelumnya, penelitian ini menggunakan regresi linear berganda untuk memprediksi harga rumah. Hasil dari penelitian ini diharapkan dapat memberikan nilai R^2 yang signifikan dan membandingkan performa metode regresi linear berganda dengan metode GRNN dalam konteks prediksi harga rumah. Studi ini juga bertujuan untuk mengevaluasi kelebihan regresi linear berganda dalam mengidentifikasi hubungan linier antara variabel prediktor dan variabel target, terutama pada data properti yang memiliki tren linier yang kuat. Dengan merujuk pada penelitian-penelitian tersebut, penelitian ini bertujuan untuk mengembangkan model regresi linear berganda yang lebih akurat serta membandingkan performanya dengan metode yang telah ada.

METODE PENELITIAN

Proses penelitian ini melibatkan tujuh tahapan yang saling berkesinambungan. Penelitian dimulai dengan identifikasi masalah untuk menentukan fokus penelitian dan tujuan yang ingin dicapai. Selanjutnya, dilakukan penentuan metode penelitian yang relevan dengan permasalahan yang dihadapi. Setelah itu, data yang diperlukan dikumpulkan melalui proses pengumpulan dataset. Data tersebut kemudian diproses pada tahap preprocessing untuk menghilangkan informasi yang tidak relevan dan memastikan kualitas data yang akan digunakan dalam analisis. Langkah berikutnya adalah penerapan metode Multiple Linear Regression untuk membangun model yang menjelaskan hubungan antara variabel independen dan variabel dependen. Setelah model terbentuk, dilakukan pengujian untuk memastikan keandalan dan validitas model yang dihasilkan. Akhirnya, penelitian diakhiri dengan penarikan kesimpulan berdasarkan hasil analisis. Tahapan-tahapan tersebut dirangkum pada Gambar 1, yang disajikan setelah penjelasan ini untuk memberikan visualisasi proses penelitian.



Gambar 1. Diagram Alur Penelitian

1. Identifikasi Masalah

Penelitian ini bertujuan untuk menganalisis pengaruh variabel umur rumah, jarak rumah ke stasiun MRT, dan jumlah toko terdekat terhadap harga rumah. Selain itu, penelitian juga mengidentifikasi hubungan keeratan antara variabel tersebut menggunakan metode regresi linier berganda [12], [13],

- [14]. Masalah utama yang dihadapi adalah ketidakpastian kontribusi setiap variabel dalam memengaruhi harga rumah, serta hubungan saling mempengaruhi antar variabel. Dalam penelitian ini, penting untuk memastikan terpenuhinya asumsi regresi linier berganda seperti linearitas, tidak adanya multikolinearitas, homoskedastisitas, dan normalitas residual agar model yang dihasilkan valid dan reliabel.
2. Metode Penelitian
Penelitian menggunakan pendekatan kuantitatif dengan metode deskriptif dan prediktif. Analisis dilakukan pada data penilaian real estate untuk memahami hubungan antar variabel. Multiple Linear Regression digunakan untuk membangun model yang mengukur pengaruh variabel independen terhadap variabel dependen.
 3. Pengumpulan Dataset
Data sekunder yang digunakan berasal dari data penilaian real estate price prediction. Pengumpulan data dengan metode ini dilakukan dengan mengunduh data berbentuk CSV yang diperoleh dari situs <https://www.kaggle.com/datasets/quantbruce/real-estate-price-prediction/data>. Data tersebut berisi beberapa kolom yaitu, transaction date, house age, distance to nearest MRT station, number of convenience stores, Latitude, Longitude, House price of unit area. Berikut adalah tabel dari data yang digunakan untuk penelitian. Data berikut akan diolah menggunakan python untuk mendapatkan hasil penelitian ini [15].
 4. Preprocessing Data
Tahap preprocessing mencakup analisis kualitas data dan penghapusan kolom yang tidak relevan untuk menghasilkan data terstruktur sesuai kebutuhan analisis.
 5. Multiple Linear Regression
Dilakukan tahapan regresi linier berganda untuk mencari nilai koefisien persamaan variabel.

$$Y = a + b_1 X_1 + b_2 X_2 + \dots + b_n X_n$$

Dengan keterangan :

- Y = Variabel Terikat.
- X1,X2,X3 = Variabel bebas
- a = Konstanta
- b1, b2, b3= Koefisien Regresi

Koefisien Determinasi (r^2)

$$r^2 = \frac{(b_1 \sum x_1) + (b_2 \sum x_2) + (b_3 \sum x_3 y)}{\sum y^3}$$

Apabila r^2 bernilai 0 , maka dalam model persamaan regresi yang terbentuk, variasi variabel tak bebas Y tidak sedikitpun dapat dijelaskan oleh variasi variabel-variabel bebas X1,X2 dan X3. Sebaliknya Apabila r^2 bernilai 1, maka dalam model persamaan regresi yang terbentuk, variabel tak bebas Y secara sempurna dapat dijelaskan oleh variasi variabel bebas X1,X2, dan X3[8], [9].

6. Pengujian Model
Pengujian dilakukan untuk memastikan asumsi klasik seperti linearitas, multikolinearitas, homoskedastisitas, dan normalitas residual terpenuhi. Pengujian ini bertujuan untuk memvalidasi model regresi yang dihasilkan[16].
7. Hasil dan Kesimpulan
Kesimpulan penelitian ditarik berdasarkan hasil analisis data, meliputi pengaruh masing-masing variabel terhadap harga rumah serta hubungan antar variabel.

HASIL PENELITIAN

Pada bagian ini, hasil penelitian akan dipaparkan berdasarkan data yang telah dikumpulkan dan dijelaskan dalam bagian metodologi penelitian. Data yang digunakan telah diperoleh melalui metode dokumentasi dan diolah menggunakan Python untuk mendapatkan hasil analisis yang relevan. Untuk memberikan gambaran lebih jelas mengenai dataset yang digunakan, tabel di bawah ini menyajikan informasi utama dari data yang dianalisis dalam penelitian ini.

Tabel 1. Dataset

No	Transaction date	House age	Distance to the nearest MRT station	Number of convenience stores	Latitude	Longitude	House price of unit area
1	2012.917	32	84.87882	10	24.98298	121.54024	37.9
2	2012.917	19.5	306.5947	9	24.98034	121.53951	42.2
3	2013.583	13.3	561.9845	5	24.98746	121.54391	47.3
4	2013.5	13.3	561.9845	5	24.98746	121.54391	54.8
5	2012.833	5	390.5684	5	24.97937	121.54245	43.1
	...	...	...	...	...	...	...
410	2013	13.7	4082.015	0	24.94155	121.50381	15.4
411	2012.667	5.6	90.45606	9	24.97433	121.5431	50
412	2013.25	18.8	390.9696	7	24.97923	121.53986	40.6
413	2013	8.1	104.8101	5	24.96674	121.54067	52.5
414	2013.5	6.5	90.45606	9	24.97433	121.5431	63.9

Preprocessing

	HouseAge	DistanceToMRT	NumberOfStores	Latitude	Longitude	HousePriceOfUnitArea
count	414.000000	414.000000	414.000000	414.000000	414.000000	414.000000
mean	17.712560	1083.885689	4.094203	24.969030	121.533361	37.980193
std	11.392485	1262.109595	2.945562	0.012410	0.015347	13.606488
min	0.000000	23.382840	0.000000	24.932070	121.473530	7.600000
25%	9.025000	289.324800	1.000000	24.963000	121.528085	27.700000
50%	16.100000	492.231300	4.000000	24.971100	121.538630	38.450000
75%	28.150000	1454.279000	6.000000	24.977455	121.543305	46.600000
max	43.800000	6488.021000	10.000000	25.014590	121.566270	117.500000

Gambar 2 Statistika Deskriptif

Berdasarkan uraian gambar 2 rata-rata harga rumah adalah 37,98 juta nilai minimum harga rumah adalah 8 juta. nilai maksimum harga rumah adalah 118 juta.

PEMBAHASAN

OLS Regression Results						
Dep. Variable:	HousePriceOfUnitArea	R-squared:	0.541			
Model:	OLS	Adj. R-squared:	0.538			
Method:	Least Squares	F-statistic:	161.1			
Date:	Wed, 24 Apr 2024	Prob (F-statistic):	5.44e-69			
Time:	12:06:12	Log-Likelihood:	-1506.5			
No. Observations:	414	AIC:	3021.			
Df Residuals:	410	BIC:	3037.			
Df Model:	3					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
const	42.9773	1.385	31.041	0.000	40.256	45.699
HouseAge	-0.2529	0.040	-6.305	0.000	-0.332	-0.174
DistanceToMRT	-0.0054	0.000	-11.874	0.000	-0.006	-0.004
NumberOfStores	1.2974	0.194	6.678	0.000	0.916	1.679
Omnibus:	222.427	Durbin-Watson:	2.115			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	3065.900			
Skew:	1.958	Prob(JB):	0.00			
Kurtosis:	15.744	Cond. No.	5.09e+03			

Gambar 3 Scatter Plot Harga Rumah dan Usia Rumah

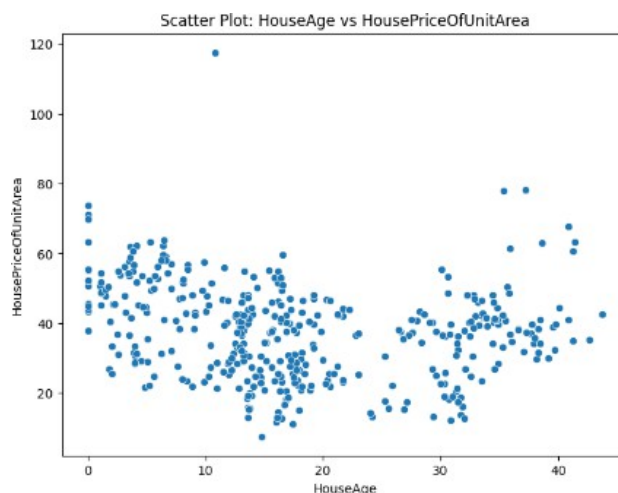
Pada hasil tabel OLS diatas didapatkan hasil nilai R-Square, F Statistik, Koefisien setiap variabel, nilai t hitung dan Durbin Watson yang akan digunakan dalam metode persamaan regresi linier berganda untuk melihat hasil persamaan dan juga hasil pengujian dari data dan model metode. Berdasarkan hasil regresi diatas didapatkan persamaan sebagai berikut.

$$y = 42,977 + (-0,253)x_1 + (-0,005)x_2 + (1,297)x_3$$

Setelah melakukan penelitian didapatkan rumus untuk mendapatkan prediksi harga rumah yang diinginkan pembeli. Rumus tersebut menjelaskan jika umur rumah bertambah 1 tahun maka harga rumah tersebut akan mengalami penurunan harga sejumlah 0,253 juta ping(dolar Taiwan), setiap penambahan jarak 1 meter dari MRT maka terjadi pengurangan harga rumah sebesar 0,005 juta ping(dolar Taiwan). dan setiap penambahan 1 jumlah toko maka akan terjadi penambahan sebesar 1,297 juta ping(dolar Taiwan).

Setelah mendapatkan rumus untuk menghitung prediksi harga rumah, berikut merupakan hasil visualisasi atau persebaran 3 faktor utama terhadap harga rumah.

- Persebaran harga rumah berdasarkan Umur Rumah

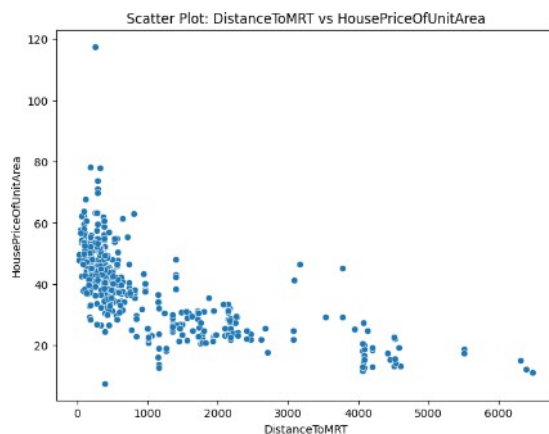


Gambar 4 Scatter Plot Harga Rumah dan Usia Rumah

Plotting pada gambar 5 menunjukkan hubungan umur rumah dengan harga rumah dengan

menggunakan plot sebar. Penyajian plot dengan atribut harga dan umur rumah bertujuan untuk memudahkan dalam menganalisis hubungan seperti berapa harga rumah dengan umur rumah, apakah semakin baru rumahnya maka semakin mahal harganya atau bukan. Gambar menunjukkan bahwa rentang umur rumah adalah dari 0 tahun hingga 40 tahun dan ada pula yang berusia lebih dari 40 tahun.

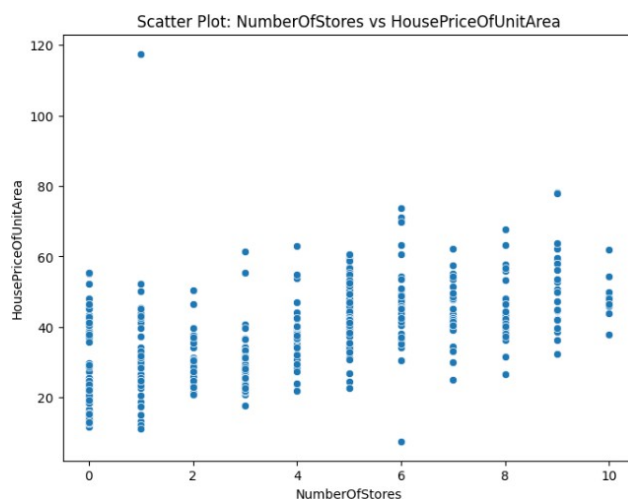
a. Persebaran harga rumah berdasarkan jarak ke MRT



Gambar 5 Scatter Plot Harga Rumah dan Jarak Rumah ke Stasiun MRT

Plot gambar 6 menggambarkan hubungan antara harga rumah (sumbu y) dan jarak rumah dari stasiun MRT (sumbu x). Titik-titik plot berkumpul pada jarak 0 hingga 1000 meter dengan dominasi warna hijau, menunjukkan bahwa sebagian besar rumah terletak dalam jarak tersebut. Hal ini mencerminkan bahwa harga rumah semakin mahal semakin dekat dengan stasiun MRT, yang menjadi pertimbangan penting bagi pembeli dalam merencanakan anggaran tambahan jika ingin membeli rumah di dekat stasiun.

b. Persebaran harga rumah berdasarkan jumlah toko terdekat



Gambar 6 Scatter Plot Harga Rumah dan Banyaknya Toko yang Dekat dengan Rumah

Plot gambar 6 menggambarkan hubungan antara harga rumah (sumbu y) dan jumlah toko yang dekat dengan rumah. Titik-titik menyebar dari 0 hingga 10 toko, dengan tren peningkatan harga seiring bertambahnya jumlah toko di sekitar rumah. Hal ini menunjukkan bahwa semakin banyak toko yang berada di dekat rumah, semakin tinggi harga rumah tersebut, yang menjadi pertimbangan penting dalam transaksi pembelian rumah. Plotting data bertujuan untuk memudahkan pemahaman pembaca mengenai penyebaran dan kondisi data. Hasil model menunjukkan nilai R-squared sebesar 54,1%, yang berarti variabel umur rumah, jarak ke stasiun MRT, dan jumlah toko terdekat mempengaruhi harga rumah sebesar 54,1%, sementara sisanya 45,9% dipengaruhi oleh faktor lain. Koefisien regresi untuk variabel umur rumah (x_1) adalah -0,253, jarak ke MRT (x_2)

-0,005, dan jumlah toko (x_3) 1,297. Persamaan regresinya adalah $y = 42,977 - 0,253x_1 - 0,005x_2 + 1,297x_3$, yang menunjukkan hubungan signifikan antara variabel, dimana perubahan satu variabel akan mempengaruhi variabel lainnya.

KESIMPULAN

Dalam analisis regresi linear ini, model menunjukkan nilai *R-squared*, yang menandakan umur rumah, jarak rumah ke mrt dan jumlah toko terdekat sekitar mempengaruhi ke harga rumah sebesar 54,1% dan sisanya 45,9% dipengaruhi oleh variabel lain. Variabel X_1 , X_2 , X_3 memiliki pengaruh signifikan terhadap harga jual rumah. Koefisien X_1 sebesar -0,253 dan koefisien X_2 sebesar -0,005 dan koefisien X_3 sebesar 1,297. Serta persamaan regresi yang dihasilkan dari data tersebut adalah sebagai berikut. $y = 42,977 - 0,253(x_1) - 0,005(x_2) + 1,297(x_3)$. Hubungan antar variabel cukup sangat terikat antara satu variabel dan variabel lainnya. Ini dibuktikan dengan kenaikan jumlah variabel maka akan mempengaruhi variabel lainnya.

REFERENCES

- [1] Undang-Undang Republik Indonesia Nomor 1 Tahun 2011 tentang Perumahan dan Kawasan Permukiman. 2011. [Online]. Available: <https://peraturan.bpk.go.id/Details/38695/uu-no-1-tahun-2011>
- [2] L. R. Fauzia, "Determinan Harga Rumah di Indonesia," *Jurnal Ekonomi dan Bisnis*, vol. 22, no. 2, pp. 45–60, 2019.
- [3] X. Fang, "Factors Influencing Housing Prices: A Case Study of Shanghai," *Theoretical and Natural Science*, vol. 52, no. 1, pp. 122–127, Sep. 2024, doi: 10.54254/2753-8818/52/2024CH0125.
- [4] E. Monica and A. Aditra, "Analisis Faktor-Faktor yang Mempengaruhi Harga Rumah di Kota Bandung," *Jurnal Ilmiah Ekonomi dan Bisnis*, vol. 10, no. 2, pp. 141–149, 2022.
- [5] L. Yan, "Predicting House Prices with a Linear Regression Model," *Journal of Applied Mathematics and Statistics*, vol. 114, pp. 1–10, 2024, doi: 10.54254/2755-2721/114/2024.18220.
- [6] M. L. Mu'tashim, S. A. Damayanti, H. N. Zaki, T. Muhayat, and R. Wirawan, "Analisis Prediksi Harga Rumah Sesuai Spesifikasi Menggunakan Multiple Linear Regression," *Jurnal Teknologi Informasi*, vol. 3, no. 1, pp. 45–56, 2021.
- [7] I. C. Yeh and T. K. Hsu, "Building Real Estate Valuation Models with Comparative Approach Through Case-Based Reasoning," *Applied Soft Computing*, vol. 65, pp. 260–271, Apr. 2018, doi: 10.1016/j.asoc.2018.01.029.
- [8] I. M. Yuliara, "Regresi Linier Berganda," *Jurnal Fisika dan Matematika*, vol. 5, no. 2, pp. 78–90, 2016.
- [9] S. Regresi and L. Berganda, "Bab 10: Regresi Linear Berganda," in *Buku Ajar Statistika Lanjutan*, 1st ed., Jakarta: Penerbit Akademik, 2020, pp. 120–135.
- [10] I. Ghozali, *Aplikasi Analisis Multivariate dengan Program IBM SPSS 25*, 9th ed. Semarang: Badan Penerbit Universitas Diponegoro, 2018. [Online]. Available: http://slims.umn.ac.id/index.php?p=show_detail&id=19545
- [11] E. F. Rahayuningtyas, F. N. Rahayu, and Y. Azhar, "Prediksi Harga Rumah Menggunakan General Regression Neural Network," *Jurnal Informatika*, vol. 8, no. 1, pp. 23–34, 2021. [Online]. Available: <https://archive.ics.uci.edu/ml/datasets/Real>
- [12] M. A. Yusuf et al., "Analisis Regresi Linier Sederhana dan Berganda Beserta Penerapannya," *Journal on Education*, vol. 6, no. 2, pp. 45–60, 2024.
- [13] W. K. O. Ho, B. S. Tang, and S. W. Wong, "Predicting Property Prices with Machine Learning Algorithms," *Journal of Property Research*, vol. 38, no. 1, pp. 48–70, 2021, doi: 10.1080/09599916.2020.1832558.
- [14] R. N. T. Siregar, V. Sitorus, and W. P. Ananta, "Analisis Prediksi Harga Rumah di Bandung Menggunakan Regresi Linear Berganda," *Journal of Creative Student Research*, vol. 1, no. 6, pp. 395–404, Dec. 2023, doi: 10.55606/jcsrpolitama.v1i6.3038.
- [15] "Real Estate Price Prediction," Kaggle, 2025. [Online]. Available:

<https://www.kaggle.com/datasets/quantbruce/real-estate-price-prediction/data>

- [16] "Memahami Uji Multikolinearitas dalam Model Regresi," Binus Accounting, 2021. [Online]. Available: <https://accounting.binus.ac.id/2021/08/06/memahami-uji-multikolinearitas-dalam-model-regresi/>