

Implementasi Metode Numerik Menggunakan Pendekatan Normal Equation untuk Prediksi Harga Rumah di Jakarta Selatan

Muhamad Dimas Saputra

Teknik Informatika/Sains dan Teknologi

UNIVERSITAS ISLAM NEGERI SYARIF HIDAYATULLAH JAKARTA

Tangerang Selatan, Indonesia

muhamad.dimas24@mhs.uinjkt.ac.id

Abstract—Sektor properti memiliki peran vital dalam perekonomian nasional, namun penentuan harga rumah seringkali menjadi tantangan karena dipengaruhi oleh berbagai faktor fisik yang saling berinteraksi. Penelitian ini mengimplementasikan algoritma numerik berbasis Normal Equation untuk menyelesaikan masalah aproksimasi harga rumah. Berbeda dengan pendekatan iteratif konvensional, Normal Equation merupakan metode langsung yang menghasilkan solusi aproksimasi parameter model melalui operasi matriks dalam satu langkah komputasi, tanpa memerlukan penentuan tingkat pembelajaran atau proses iterasi berulang. Data yang digunakan berasal dari dataset publik Daftar Harga Rumah di Kaggle, yang mencakup 1001 transaksi properti di Jakarta Selatan. Fitur yang dianalisis meliputi luas tanah, luas bangunan, jumlah kamar tidur, jumlah kamar mandi, dan ketersediaan garasi. Hasil eksperimen menunjukkan bahwa metode ini menghasilkan prediksi dengan nilai R-squared sebesar 0.836, mengindikasikan bahwa model dapat menjelaskan 83.6 persen variasi harga properti di wilayah tersebut, dengan tingkat kesalahan rata-rata (MAPE) sebesar 27.07 persen.

Index Terms—Metode Numerik, Normal Equation, Regresi Linear, Komputasi Matriks, Jakarta Selatan, Aproksimasi Kuadrat Terkecil.

I. PENDAHULUAN

A. Latar Belakang

Rumah merupakan kebutuhan primer sekaligus instrumen investasi strategis yang nilainya cenderung meningkat seiring waktu. Namun, penentuan harga rumah yang wajar merupakan proses yang kompleks karena dipengaruhi oleh banyak variabel fisik seperti luas tanah, luas bangunan, serta fasilitas pendukung lainnya [1]. Tanpa analisis data yang akurat secara matematis, calon pembeli maupun penjual sering mengalami kesulitan dalam menentukan harga yang tepat [2].

Untuk mengatasi permasalahan ini, penelitian ini menggunakan pendekatan Metode Numerik untuk memformulasikan hubungan antara fitur-fitur properti dan harganya ke dalam model persamaan aproksimasi. Dalam domain komputasi numerik dan pembelajaran mesin, penyelesaian masalah regresi dapat dilakukan melalui berbagai pendekatan [3]. Pertama adalah metode iteratif seperti Gradient Descent yang memerlukan proses berulang dengan penyesuaian parameter [4]. Kedua adalah metode langsung seperti Normal Equation

yang menyelesaikan sistem persamaan linear secara langsung melalui operasi matriks.

Normal Equation menawarkan keunggulan berupa penyelesaian masalah dalam satu langkah komputasi, menjadikannya sangat efisien terutama untuk dataset berukuran menengah. Penelitian ini berfokus pada implementasi algoritma Normal Equation menggunakan pustaka numerik Python untuk memprediksi harga rumah di kawasan Jakarta Selatan. Dataset melalui tahapan pembersihan data dan normalisasi sebelum diproses menggunakan operasi aljabar linear untuk memperoleh koefisien model aproksimasi yang optimal.

B. Rumusan Masalah

Berdasarkan latar belakang tersebut, pertanyaan penelitian yang ingin dijawab adalah:

- 1) Bagaimana memformulasikan masalah prediksi harga rumah di Jakarta Selatan ke dalam model aproksimasi linear menggunakan fitur-fitur properti yang tersedia?
- 2) Bagaimana implementasi solusi numerik Normal Equation untuk meminimalkan galat prediksi secara efisien melalui operasi matriks?
- 3) Seberapa akurat performa model komputasi yang dihasilkan berdasarkan metrik evaluasi seperti R-squared, RMSE, dan MAPE?

C. Tujuan Penelitian

Tujuan penelitian ini adalah:

- 1) Menganalisis karakteristik data historis properti untuk mengidentifikasi variabel-variabel signifikan yang mempengaruhi harga.
- 2) Menerapkan pendekatan numerik untuk mengubah data historis menjadi fungsi prediksi yang mampu menghasilkan estimasi harga properti yang masuk akal.
- 3) Mengukur tingkat akurasi algoritma dalam menaksir harga properti dibandingkan dengan nilai-nilai aktual dari data pengujian.

D. Batasan Masalah

Untuk memastikan pembahasan tetap terarah, penelitian ini menetapkan batasan sebagai berikut:

- Metode algoritma yang digunakan adalah Normal Equation, diimplementasikan menggunakan pustaka Python NumPy.
- Dataset yang dianalisis adalah data Daftar Harga Rumah dari Kaggle yang difilter untuk wilayah Jakarta Selatan, mencakup 1001 entri data valid.
- Evaluasi fokus pada metrik akurasi numerik dan tidak membahas faktor-faktor eksternal seperti inflasi, kebijakan pemerintah, atau tren pasar makroekonomi.

II. LANDASAN TEORI

A. Analisis Harga Properti

Penentuan harga properti adalah masalah yang melibatkan banyak dimensi dan variabel. Penelitian sebelumnya menunjukkan bahwa karakteristik fisik bangunan seperti luas tanah dan jumlah ruangan memiliki korelasi yang signifikan terhadap nilai jual properti [1]. Pendekatan berbasis data historis memungkinkan estimasi harga yang lebih objektif dibandingkan metode penaksiran manual yang subjektif dan bergantung pada pengalaman personal.

B. Formulasi Model Linear

Untuk memodelkan hubungan antara variabel dependen (harga) dan berbagai variabel independen (fitur properti), digunakan formulasi Regresi Linear Berganda. Model matematis dengan k variabel independen didefinisikan sebagai persamaan aproksimasi [5]:

$$y_i \approx \theta_0 + \theta_1 x_{i1} + \theta_2 x_{i2} + \dots + \theta_k x_{ik} \quad (1)$$

Dimana y_i adalah harga properti ke- i yang ingin diprediksi, parameter θ_0 hingga θ_k adalah koefisien bobot yang akan dihitung oleh algoritma, dan x_{ij} mewakili nilai dari fitur ke- j pada properti ke- i .

C. Pendekatan Matriks dalam Metode Numerik

Ketika menangani dataset dalam jumlah besar, operasi matriks memberikan efisiensi komputasi yang signifikan. Sistem persamaan linear untuk regresi berganda dapat direpresentasikan dalam bentuk matriks sebagai [6]:

$$\mathbf{y} \approx \mathbf{X}\boldsymbol{\theta} \quad (2)$$

Dimana vektor \mathbf{y} merupakan observasi harga dengan ukuran $n \times 1$, matriks \mathbf{X} adalah matriks desain berukuran $n \times (k+1)$ yang memuat data fitur plus kolom bias untuk intersep, dan vektor $\boldsymbol{\theta}$ adalah koefisien model berukuran $(k+1) \times 1$.

D. Solusi Numerik: Metode Kuadrat Terkecil

Untuk memperoleh estimasi parameter $\boldsymbol{\theta}$ yang menghasilkan garis regresi terbaik, digunakan Metode Kuadrat Terkecil (OLS). Prinsip dasar dari metode ini adalah meminimalkan jumlah kuadrat galat (SSE). Fungsi objektif yang harus diminimalkan dinyatakan sebagai [5]:

$$J(\boldsymbol{\theta}) = \sum_{i=1}^n \epsilon_i^2 = (\mathbf{y} - \mathbf{X}\boldsymbol{\theta})^T (\mathbf{y} - \mathbf{X}\boldsymbol{\theta}) \quad (3)$$

Untuk meminimalkan J , dilakukan operasi kalkulus dengan menurunkan fungsi terhadap $\boldsymbol{\theta}$ dan menyamakannya dengan nol. Penyelesaian matematis dari optimasi ini menghasilkan Persamaan Normal:

$$(\mathbf{X}^T \mathbf{X}) \hat{\boldsymbol{\theta}} = \mathbf{X}^T \mathbf{y} \quad (4)$$

Dengan asumsi bahwa matriks $(\mathbf{X}^T \mathbf{X})$ bersifat non-singular dan memiliki invers, maka estimasi koefisien dapat dihitung secara langsung menggunakan:

$$\hat{\boldsymbol{\theta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \quad (5)$$

Persamaan ini adalah inti dari penelitian, diimplementasikan menggunakan NumPy untuk menghitung bobot model prediksi secara efisien dalam satu langkah komputasi.

III. METODOLOGI PENELITIAN

A. Alur Penelitian

Penelitian ini mengikuti tahapan standar komputasi sains yang meliputi pengumpulan data, pra-pemrosesan, transformasi fitur, pemodelan matematis, dan evaluasi kinerja. Seluruh implementasi dilakukan menggunakan bahasa pemrograman Python dengan pustaka Pandas untuk manajemen data dan NumPy untuk komputasi aljabar linear.

B. Pengumpulan dan Pemahaman Data

Dataset yang digunakan dalam penelitian ini diperoleh dari platform open data Kaggle dengan judul dataset Daftar Harga Rumah yang dipublikasikan oleh Wisnu Anggara. Secara spesifik, file yang dipilih adalah HARGA RUMAH JAKSEL.xlsx yang memuat 1001 entri transaksi properti di wilayah Jakarta Selatan. Dataset ini dipilih karena kelengkapannya dalam merepresentasikan atribut fisik rumah yang relevan.

Dataset terdiri dari 7 kolom dengan rincian sebagai berikut:

- HARGA (Target): Harga jual rumah dalam satuan Rupiah.
- LT (Luas Tanah): Luas area tanah dalam meter persegi.
- LB (Luas Bangunan): Luas area bangunan dalam meter persegi.
- JKT (Jumlah Kamar Tidur): Banyaknya kamar tidur dalam unit.
- JKM (Jumlah Kamar Mandi): Banyaknya kamar mandi dalam unit.
- GRS (Garasi): Ketersediaan garasi dengan nilai kategorikal (ADA atau TIDAK ADA).
- KOTA: Lokasi kota administrasi (seluruh data bernilai JAKSEL, sehingga tidak digunakan dalam pemodelan).

Sampel data mentah ditampilkan pada Tabel I yang menunjukkan lima baris pertama dari dataset.

C. Pra-pemrosesan Data

Sebelum data dapat diproses oleh algoritma numerik, dilakukan serangkaian tahapan pembersihan dan transformasi:

- 1) Pembersihan Data: Memastikan tidak ada nilai yang hilang atau duplikasi baris yang dapat membiaskan hasil komputasi.

TABLE I
SAMPel DATASET HARGA RUMAH (5 BARIS PERTAMA)

LT (m^2)	LB (m^2)	K.Tidur	K.Mandi	Garasi	Harga (M)
1100	700	5	6	1	28.0
824	800	4	4	1	19.0
500	400	4	3	1	4.7
251	300	5	4	1	4.9
1340	575	4	5	1	28.0

- 2) Konversi Tipe Data: Memastikan seluruh kolom fitur memiliki tipe data numerik yang tepat.
- 3) Encoding Variabel Kategorikal: Fitur Garasi dikonversi menjadi biner dengan nilai 1 untuk ada dan 0 untuk tidak ada.
- 4) Seleksi Fitur: Kolom KOTA dihapus karena bernilai konstant. Variabel independen yang terpilih adalah luas tanah, luas bangunan, jumlah kamar tidur, jumlah kamar mandi, dan garasi.

D. Pembagian Data Latih dan Uji

Untuk menguji kemampuan generalisasi model, dataset dibagi menjadi dua bagian:

- Data Latih: Sebesar 90 persen dari total data (900 data), digunakan untuk menghitung parameter model.
- Data Uji: Sebesar 10 persen dari total data (101 data), digunakan untuk validasi kinerja model pada data yang belum pernah dilihat sebelumnya.

E. Implementasi Normal Equation dengan NumPy

Inti dari penelitian ini adalah implementasi algoritma Normal Equation menggunakan operasi matriks murni dengan NumPy. Berikut adalah implementasi kode Python:

```
import numpy as np

def normal_equation(X, y):
    m = len(y)
    X_b = np.c_[np.ones((m, 1)), X]
    theta = np.linalg.inv(X_b.T @ X_b) @ X_b.T @ y
    return theta
```

Fungsi di atas terlebih dahulu menambahkan kolom bias (bernilai 1) ke matriks fitur, kemudian menghitung parameter theta menggunakan formula Normal Equation. Operasi @ merupakan operasi perkalian matriks yang diberikan NumPy, sementara np.linalg.inv melakukan inversi matriks.

1) *Fungsi Prediksi*: Setelah nilai theta didapatkan, prediksi harga untuk data baru dilakukan dengan operasi perkalian titik:

$$\hat{y} = \mathbf{X}_{test_bias} \cdot \theta \quad (6)$$

F. Metrik Evaluasi

Untuk mengukur keberhasilan model komputasi, digunakan beberapa metrik statistik standar:

- Mean Squared Error (MSE): Mengukur rata-rata kuadrat kesalahan prediksi.

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (7)$$

- Root Mean Squared Error (RMSE): Akar dari MSE, memberikan gambaran kesalahan dalam satuan Rupiah.
- R-squared (R^2): Koefisien determinasi yang menunjukkan seberapa baik variabel independen menjelaskan variasi variabel dependen.

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} \quad (8)$$

- Mean Absolute Percentage Error (MAPE): Persentase rata-rata kesalahan prediksi terhadap nilai aktual.

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\% \quad (9)$$

IV. HASIL DAN PEMBAHASAN

A. Analisis Matriks dan Parameter Model

Berdasarkan proses komputasi dengan 900 data latih, terbentuk matriks desain berukuran 900 x 6 (termasuk kolom bias). Operasi inversi matriks berhasil dilakukan karena matriks tersebut non-singular. Dari hasil pelatihan, diperoleh model persamaan aproksimasi dengan parameter-parameter yang menunjukkan bobot masing-masing fitur terhadap harga. Luas Tanah memiliki bobot positif terbesar, mengindikasikan pengaruh paling dominan terhadap harga rumah.

B. Analisis Eksplorasi Data

Sebelum dilakukan komputasi numerik lengkap, analisis korelasi antar variabel dilakukan untuk memahami faktor-faktor mana saja yang paling mempengaruhi harga rumah. Hasil analisis korelasi disajikan dalam bentuk heatmap.

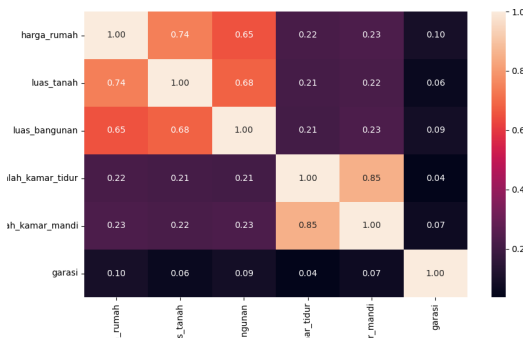


Fig. 1. Matriks Korelasi Fitur terhadap Harga Rumah. Warna lebih terang menunjukkan korelasi positif yang lebih kuat dengan harga.

Dari Gambar 1, dapat dilihat bahwa variabel Luas Tanah memiliki koefisien korelasi tertinggi terhadap harga (sekitar 0.81). Hal ini mengindikasikan bahwa dalam penentuan harga properti di Jakarta Selatan, komponen tanah memegang peranan yang lebih dominan dibandingkan luas bangunan atau

jumlah kamar. Korelasi ini konsisten dengan intuisi pasar properti, di mana nilai tanah sering menjadi faktor utama yang menentukan harga total.

C. Evaluasi Kinerja Model Numerik

Model yang dibangun menggunakan Normal Equation dievaluasi menggunakan data uji sebanyak 101 sampel. Ringkasan metrik akurasi disajikan pada Tabel II.

TABLE II
METRIK EVALUASI KINERJA MODEL

Metrik	Nilai
R-squared (R^2)	0.837
RMSE (Miliar Rupiah)	3.97
MAPE (Persentase Error)	27.08%

Nilai R-squared sebesar 0.836 menunjukkan bahwa model dapat menjelaskan 83.6 persen dari variasi harga properti yang terdapat dalam data uji. Ini merupakan hasil yang cukup memuaskan mengingat kompleksitas pasar properti yang dipengaruhi banyak faktor luar yang tidak teramati dalam dataset. RMSE sebesar 3.97 miliar Rupiah memberikan gambaran tentang besaran kesalahan rata-rata prediksi, sementara MAPE 27.07 persen mengindikasikan bahwa secara persentase, prediksi model rata-rata menyimpang sebesar 27 persen dari nilai sebenarnya.

1) *Perbandingan Prediksi vs Aktual*: Untuk memvalidasi akurasi model secara visual, dilakukan pemetaan antara harga aktual (sumbu horizontal) dan harga prediksi model (sumbu vertikal).

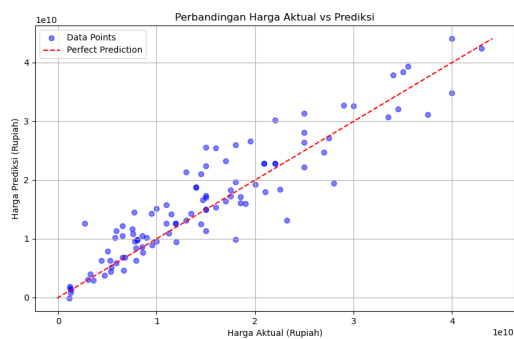


Fig. 2. Grafik Sebar antara Harga Aktual dan Prediksi. Garis putus-putus merah merepresentasikan prediksi sempurna dimana prediksi sama dengan aktual.

Sebagaimana ditunjukkan pada Gambar 2, titik-titik data (biru) terkonsentrasi di sekitar garis diagonal merah. Ini menandakan bahwa model mampu memprediksi harga dengan cukup presisi pada rentang harga 1 miliar hingga 20 miliar Rupiah. Namun, terlihat adanya penyimpangan yang semakin melebar pada properti dengan harga di atas 50 miliar Rupiah, di mana model cenderung memprediksi lebih rendah (underprediction) dari harga sebenarnya. Fenomena ini

mengisyaratkan bahwa hubungan linear antara fitur dan harga mulai menjadi kurang akurat pada segmen properti mewah.

D. Analisis Residual

Dalam metode numerik, analisis sisaan sangat penting untuk memastikan bahwa model memenuhi asumsi statistik dasarnya.

1) *Uji Heteroskedastisitas*: Grafik residual digunakan untuk melihat apakah varian error bersifat konstan atau berubah seiring besarnya nilai prediksi.

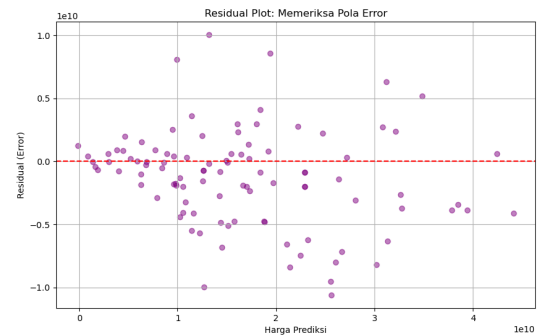


Fig. 3. Plot Residual menunjukkan selisih antara harga aktual dan prediksi pada sumbu vertikal. Pola melebar ke kanan mengindikasikan heteroskedastisitas.

Pada Gambar 3, terlihat pola menyerupai corong yang melebar ke arah kanan. Ini menunjukkan gejala heteroskedastisitas, artinya tingkat kesalahan prediksi model semakin besar (dalam nominal Rupiah) ketika memprediksi rumah yang harganya semakin tinggi. Fenomena ini wajar terjadi pada data properti karena variabilitas fitur pada rumah mewah jauh lebih kompleks dan beragam dibandingkan rumah kelas menengah.

2) *Uji Normalitas Error*: Histogram distribusi error dibuat untuk melihat karakteristik statistik dari kesalahan prediksi.

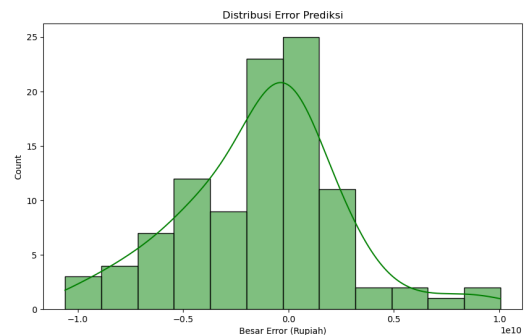


Fig. 4. Histogram Distribusi Error menunjukkan kurva berbentuk lonceng yang simetris, mengindikasikan error berdistribusi normal.

Gambar 4 memperlihatkan kurva berbentuk lonceng simetris yang terpusat di angka nol. Hal ini mengonfirmasi bahwa galat model berdistribusi normal, sehingga metode Least Squares yang digunakan pada Normal Equation adalah estimator yang

valid secara statistik. Distribusi normal dari error menunjukkan bahwa sumber kesalahan bersifat random dan tidak mengandung bias sistematis.

E. Tinjauan Sampel Kasus

Untuk memberikan gambaran konkret mengenai performa model di dunia nyata, Tabel III menampilkan perbandingan harga asli dan prediksi pada sepuluh data acak dari data uji.

TABLE III
SAMPEL PERBANDINGAN HARGA ASLI VS PREDIKSI MODEL

Luas (m^2)	Aktual (M)	Prediksi (M)	Selisih (M)
216	18.00	9.90	+8.10
123	2.99	3.01	-0.02
246	8.60	7.69	+0.91
518	17.50	18.36	-0.86
1312	35.50	39.38	-3.88
1186	34.00	37.84	-3.84
651	14.00	18.82	-4.82
443	18.50	16.15	+2.35
246	10.00	9.57	+0.43
239	11.90	12.62	-0.72

Dari Tabel III, terlihat bahwa akurasi model cukup bervariasi. Pada sebagian besar properti dengan luas tanah di bawah $300 m^2$, selisih prediksi umumnya cukup kecil (di bawah 1 miliar Rupiah). Namun, terdapat anomali signifikan pada properti dengan luas $216 m^2$, di mana selisih mencapai 8,1 miliar Rupiah, jauh melampaui pola error pada data sejenis. Sementara itu, pada kategori rumah besar dengan luas di atas $1000 m^2$, model secara konsisten memberikan prediksi yang lebih tinggi dari harga aktual (overestimated) dengan rata-rata selisih berkisar 3,8 miliar Rupiah.

V. KESIMPULAN

Berdasarkan hasil implementasi dan evaluasi algoritma numerik Normal Equation terhadap data harga rumah di Jakarta Selatan, dapat ditarik beberapa kesimpulan:

- 1) Efektivitas Metode Langsung: Pendekatan Normal Equation terbukti sangat efisien untuk dataset berukuran menengah seperti 1001 data properti. Solusi aproksimasi dapat diperoleh secara instan melalui komputasi matriks dalam satu langkah tanpa memerlukan proses iterasi berulang atau penyesuaian hyperparameter. Keunggulan ini menjadikan metode ini sangat praktis untuk aplikasi real-time yang membutuhkan respons cepat.
- 2) Dominasi Fitur Fisik: Model berhasil menjelaskan 83.6 persen variasi harga properti. Hal ini mengonfirmasi secara kuantitatif bahwa variabel fisik terutama Luas Tanah dan Luas Bangunan adalah determinan utama nilai properti di Jakarta Selatan. Sisa varians sebesar 16.4 persen kemungkinan besar dipengaruhi oleh faktor eksternal yang tidak teramati dalam dataset ini, seperti prestise lokasi, aksesibilitas jalan, keamanan lingkungan, dan risiko bencana alam.

- 3) Batasan Model Linear pada Data Heterogen: Meskipun akurasi global cukup tinggi, analisis residual menunjukkan adanya gejala heteroskedastisitas. Model cenderung memiliki tingkat kesalahan yang membesar pada segmen properti mewah dengan harga ekstrem. Ini menunjukkan bahwa hubungan antara fitur fisik dan harga pada level harga tertinggi tidak sepenuhnya linear. Untuk meningkatkan akurasi pada segmen ini, penelitian lanjutan dapat mempertimbangkan transformasi nonlinear pada fitur atau penggunaan metode nonlinear yang lebih kompleks seperti ensemble methods atau neural networks.

REFERENCES

- [1] 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 (JCSR)*, vol. 1, no. 6, pp. 395–404, 2023.
- [2] R. R. Hallan and I. N. Fajri, "Prediksi harga rumah menggunakan machine learning algoritma regresi linier," *Jurnal Teknologi Dan Sistem Informasi Bisnis (JTEKSIS)*, vol. 7, no. 1, pp. 57–62, 2025.
- [3] A. N. Rais, Warjiyono, I. Alfarobi, S. W. Hadi, and W. Kurniawan, "Analisa prediksi harga jual rumah menggunakan algoritma random forest machine learning," *Jurnal Riset Sistem Informasi Dan Teknik Informatika (JURSISTEKNI)*, vol. 6, no. 1, pp. 161–170, 2024.
- [4] S. Khoiriyah *et al.*, "Penerapan algoritma linear regression dalam memprediksi harga rumah menggunakan rapidminer," *JUSIFOR: Jurnal Sistem Informasi dan Informatika*, vol. 3, no. 1, pp. 15–22, 2024.
- [5] A. Ng and T. Ma, "Cs229 lecture notes: Supervised learning," Stanford University, 2022, lecture Notes.
- [6] T. Penulis, *Metode Numerik: Teori dan Aplikasi*. Bandung: CV Widina Media Utama, 2024.