



Prediksi Harga Tanah menggunakan *Machine Learning* Algoritma Regresi Linier

Arjunnajah Muhammad Samhan^{1*}, Aprilisa Arum Sari², Pramono³

¹⁻³ Teknik Informatika, Universitas Duta Bangsa, Indonesia

*Penulis Korespondensi: arjunnajah.samhan239@gmail.com

Abstract. Conventional land price determination in Surakarta City remains manual, subjective, relying on personal assessments that tend to follow land prices in the surrounding area, and is inefficient for large-scale data. This study aims to implement the Multiple Linear Regression algorithm to predict land prices based on physical and environmental characteristics. The data used comprises 1,923 offer and transaction records for land from 2021 to 2024 sourced from the Surakarta City Land Office. The methodology employed is CRISP-DM, which consists of six stages: business understanding, data understanding, data preparation (including cleaning, normalization, and data splitting into 80% training data and 20% testing data), modeling, evaluation, and deployment. The features used include land area, road width, distance to the city center (derived from X,Y coordinates), building area, number of floors, and road class (one-hot encoding). The results show that the model achieved an $R^2 = 0.7767$, RMSE = IDR 684.3 million, and MAE = IDR 503.9 million. The RMSE is smaller than the standard deviation of land prices, IDR 1,493,000,000, indicating that the model is considered good. The model coefficients indicate that building area and land area have the largest positive influence, while distance to the city center has a negative effect. A Streamlit-based prediction prototype has been deployed to facilitate users in inputting land parameters and obtaining real-time price estimates. This study contributes an objective, fast, and easily interpretable prediction model for use by government agencies and property practitioners in decision-making.

Keywords: CRISP-DM; Land Price Prediction; Linear Regression; Machine Learning; Streamlit.

Abstrak. Penentuan harga tanah secara konvensional di Kota Surakarta masih bersifat manual, subjektif penilaian pribadi yang cenderung mengikuti harga tanah di sekitar lokasi, dan tidak efisien untuk data berskala besar. Penelitian ini bertujuan mengimplementasikan algoritma Regresi Linier Berganda untuk memprediksi harga tanah berdasarkan karakteristik fisik dan lingkungan. Data yang digunakan adalah 1.923 catatan penawaran dan transaksi tanah tahun 2021–2024 dari Kantor Pertanahan Kota Surakarta. Metodologi yang digunakan adalah CRISP-DM yang terdiri dari enam tahapan: pemahaman bisnis, pemahaman data, persiapan data mulai dari pembersihan, normalisasi, dan split data menjadi 80% data latih dan 20% data uji, pemodelan, evaluasi, dan deployment. Fitur yang digunakan meliputi luas tanah, lebar jalan, jarak ke pusat kota yang berasal dari turunan koordinat X,Y, luas bangunan, jumlah lantai, dan kelas jalan (*one-hot encoding*). Hasil penelitian menunjukkan model mencapai $R^2 = 0,7767$, RMSE = Rp 684,3 juta, dan MAE = Rp 503,9 juta. Nilai RMSE lebih kecil dari standar deviasi harga tanah Rp 1.493.000.000 sehingga model dinyatakan baik. Koefisien model menunjukkan bahwa luas bangunan dan luas tanah memiliki pengaruh positif terbesar, sementara jarak ke pusat kota berpengaruh negatif. *Prototype* prediksi berbasis *Streamlit* telah dideploy untuk memudahkan pengguna memasukkan parameter tanah dan mendapatkan estimasi harga secara real-time. Penelitian ini memberikan kontribusi berupa model prediksi yang objektif, cepat, dan mudah diinterpretasikan di instansi pemerintah dan praktisi properti dalam pengambilan keputusan.

Kata Kunci: CRISP-DM; *Machine Learning*; Prediksi Harga Tanah; Regresi Linier; *Streamlit*.

1. LATAR BELAKANG

Secara ideal, penentuan harga tanah harus dilakukan secara objektif, transparan, dan berbasis data yang akurat, sebagaimana diamanatkan dalam Peraturan Menteri ATR/BPN Nomor 3 Tahun 2026 serta Standar Penilaian Indonesia yang mengutamakan estimasi nilai pasar berdasarkan data riil (Pakusadewo et al., 2023).

Namun, realitas di lapangan menunjukkan bahwa praktik penilaian harga tanah di Kota Surakarta masih dilakukan secara konvensional dan manual, dengan pendekatan perbandingan pasar yang sangat bergantung pada penilaian subjektif serta cenderung mengikuti harga di sekitar lokasi(Yulianto, 2024.). Kondisi ini menjadi tidak efisien ketika dihadapkan pada volume data transaksi yang terus meningkat, serta menyulitkan analisis berskala besar (Risang Narindra et al., 2024). Berdasarkan dataset transaksi tanah tahun 2021–2024 dari Kantor Pertanahan Kota Surakarta, terlihat variasi harga yang cukup tinggi antara Rp200.000 hingga Rp3.500.000 per meter persegi, mengindikasikan bahwa nilai tanah dipengaruhi oleh beragam faktor fisik dan lingkungan seperti luas tanah, akses jalan, dan lokasi spasial yang kompleks(Hayya' et al., 2025).

Permasalahan utama yang timbul adalah lamanya waktu analisis, tingginya risiko subjektivitas, serta keterbatasan metode manual dalam menangani data dalam jumlah besar secara cepat dan terukur(Septriasnu, 2024). Perkembangan teknologi *Machine Learning* menawarkan peluang untuk melakukan prediksi harga tanah secara otomatis dan objektif (Akbar Marhananda & Ezra Fabiandika, 2026). Meskipun penelitian sebelumnya telah banyak menerapkan algoritma kompleks seperti *Random Forest* dan *Neural Network* untuk prediksi properti (Salman et al., 2024) penerapan algoritma Regresi Linier yang sederhana, mudah diinterpretasikan, dan efisien secara komputasi masih relatif terbatas pada konteks prediksi harga tanah di Indonesia (Hallan & Fajri, 2025). Penelitian oleh (Lathifah & Danar Dana, 2024) berhasil menerapkan Regresi Linier pada properti real estate, namun dengan variabel dan lokasi yang berbeda, sedangkan (Singgalen, 2023) lebih menekankan efektivitas kerangka CRISP-DM pada kasus klasifikasi teks, bukan pada regresi nilai tanah.

Penawaran kabaharuan terletak pada integrasi variabel spasial (koordinat X,Y , yang di gunakan untuk menghitung jarak ke pusat kota) serta karakteristik fisik tanah di Kota Surakarta, yang dikombinasikan dengan penerapan enam tahapan CRISP-DM secara sistematis hingga tahap *deployment* menggunakan *Streamlit*(Syafarina, 2023). Kegiatan ini bertujuan mengimplementasikan algoritma Regresi Linier Berganda untuk memprediksi harga tanah berdasarkan fitur luas tanah, lebar jalan, jarak ke pusat kota, luas bangunan, jumlah lantai, dan kelas jalan, serta mengidentifikasi faktor yang paling berpengaruh(Alim, 2024). Solusi yang ditawarkan adalah model prediksi yang objektif, cepat, dan mudah diinterpretasikan, yang diharapkan dapat menjadi alternatif pendukung keputusan bagi instansi pemerintah (seperti BPN) dan praktisi properti dalam menentukan estimasi nilai tanah secara lebih terukur dan berbasis data (Nurani et al., 2023).

2. KAJIAN TEORITIS

Nilai tanah didefinisikan sebagai estimasi ekonomi dari suatu bidang tanah atau zona yang mencerminkan harga wajar berdasarkan kondisi pasar pada waktu tertentu (Roustaei, 2024), dan secara yuridis formal Peraturan Menteri ATR/BPN Nomor 3 Tahun 2026 menyatakan bahwa Nilai Tanah adalah nilai dari permukaan bumi dalam batas tertentu. Nilai ini bersifat dinamis dan dipengaruhi oleh berbagai faktor saling terkait, seperti faktor ekonomi, sosial, serta faktor fisik dan infrastruktur seperti aksesibilitas, lebar jalan, luas tanah, dan kedekatan dengan fasilitas umum (Risang Narindra et al., 2024). Berfokus pada prediksi berbasis data kuantitatif, variabel independen yang digunakan adalah karakteristik fisik dan spasial yang dapat diukur secara objektif, yaitu luas tanah, lebar jalan akses, koordinat lokasi (X,Y), luas bangunan, jumlah lantai, dan kelas jalan(Khoiriyah & Fatah, 2024).

Standar operasional prosedur penilaian tanah di Indonesia diatur dalam Peraturan Menteri ATR/BPN Nomor 3 Tahun 2026, yang mengamanatkan bahwa penyelenggaraan penilaian tanah dilakukan melalui pembuatan dan pembaruan Peta Zona Nilai Tanah (ZNT) dan Peta Nilai Bidang Tanah (NBT) (Pasal 4) dengan menggunakan data transaksi dan data penawaran (Pasal 7) yang dikumpulkan melalui survei kepada responden terpilih (Pasal 10). Secara teknis, pembuatan ZNT mencakup tahapan persiapan, survei batas, pengumpulan data sampel (minimal 3 sampel per zona), serta analisis data yang meliputi penyesuaian harga penawaran menjadi transaksi, penyesuaian jenis hak menjadi Hak Milik, dan penyesuaian waktu ke tanggal penilaian (Pasal 13 dan 19). Yang perlu digarisbawahi, dalam pembuatan NBT, peraturan ini secara eksplisit mengizinkan penggunaan teknik statistika modern dan kecerdasan buatan/*machine learning* seperti *Random Forest* (Pasal 29), yang menunjukkan bahwa kebijakan nasional telah mengarah pada transformasi digital dalam penilaian tanah(Sari et al., 2024).

Adapun algoritma yang digunakan dalam penelitian ini adalah Regresi Linier Berganda, yaitu metode statistik yang memodelkan hubungan linear antara dua atau lebih variabel independen dengan variabel dependen, dengan bentuk umum persamaan $Y = a + b_1X_1 + b_2X_2 + \dots + b_nX_n + e$, di mana Y adalah harga tanah, a adalah konstanta, b adalah koefisien regresi, X adalah variabel prediktor, dan e adalah error term (Lathifah & Danar Dana, 2024). Metode ini dipilih karena model yang dihasilkan bersifat interpretable, artinya pengaruh setiap variabel terhadap harga tanah dapat dijelaskan secara kuantitatif, sehingga mudah dipahami oleh pemangku kepentingan(Hakim et al., 2025).

Evaluasi kinerja model prediksi dilakukan dengan membandingkan nilai prediksi terhadap nilai aktual menggunakan tiga metrik utama, yaitu RMSE (*Root Mean Square Error*) yang mengukur akar rata-rata kuadrat selisih prediksi dan aktual dan sensitif terhadap outlier, MAE (*Mean Absolute Error*) yang mengukur rata-rata selisih absolut dan lebih robust terhadap outlier, serta koefisien determinasi R^2 yang mengukur proporsi variasi variabel dependen yang dapat dijelaskan oleh model (Alnuaimi & Albaldawi, 2024). Karena RMSE dan MAE bersifat absolut dan bergantung pada skala harga, kriteria keberhasilan model dalam penelitian ini mengacu pada penelitian sebelumnya, yaitu model dinyatakan baik apabila RMSE lebih kecil dari standar deviasi harga tanah dan $R^2 > 0,7$ (Lathifah & Danar Dana, 2024; Putri & Arianto, 2024);. Perlu ditekankan bahwa meskipun Peraturan Menteri ATR/BPN Nomor 3 Tahun 2026 memperkenalkan Simpangan Baku Relatif (SBR) untuk menguji homogenitas sampel dalam pembuatan ZNT (Pasal 20), penelitian ini tidak menggunakan SBR sebagai metrik evaluasi karena berada pada ranah prediksi nilai bidang tanah (NBT) menggunakan *machine learning* yang diakomodir oleh Pasal 29 ayat (1), sehingga metrik yang digunakan mengacu pada standar evaluasi model prediktif di bidang *data science* yaitu RMSE, MAE, dan R^2 (Parhusip et al., 2025).

3. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan kerangka CRISP-DM yang terdiri dari enam tahapan: pemahaman bisnis, pemahaman data, persiapan data, pemodelan, evaluasi, dan *deployment*. Kerangka ini dipilih karena menyediakan alur sistematis dari data mentah hingga model prediksi yang siap digunakan. Penelitian dilatarbelakangi oleh permasalahan penentuan harga tanah di Kota Surakarta yang masih konvensional, manual, subjektif, dan tidak efisien untuk data berskala besar, dengan tujuan membangun model prediksi harga tanah yang objektif, cepat, dan akurat, serta solusi berupa model machine learning dengan antarmuka sederhana untuk estimasi harga secara real-time.

Data yang digunakan merupakan data primer berupa catatan penawaran dan transaksi tanah di Kota Surakarta tahun 2021–2024 yang diperoleh dari Kantor Pertanahan Kota Surakarta, dengan total 1.923 baris data dan 74 kolom variabel. Adapun variable yang digunakan hanya sejumlah 7 variabel utama yang tersedia antara lain luas tanah, lebar jalan akses, koordinat X dan Y, harga tanah, alamat, kelurahan, kecamatan, luas bangunan, jumlah lantai, serta berbagai atribut pendukung lainnya.

Variabel target dalam penelitian adalah harga tanah, sedangkan variabel prediktor yang digunakan meliputi luas tanah, lebar jalan, jarak ke pusat kota yang dihitung dari koordinat X dan Y, luas bangunan, jumlah lantai, serta kelas jalan yang telah melalui proses one-hot encoding.

Sebelum pemodelan, data melalui proses pembersihan dan seleksi fitur. Penanganan nilai hilang dilakukan dengan imputasi median pada variabel numerik dan penghapusan baris dengan nilai hilang pada target, sedangkan outlier dideteksi dan ditangani menggunakan metode Interquartile Range (IQR) untuk menghindari distorsi model. Seleksi fitur dilakukan dengan memilih variabel yang relevan dan berkorelasi signifikan terhadap harga tanah. Setelah dilakukannya pembersihan data, tahap selanjutnya yaitu encoding dan normalisa data.

Algoritma yang digunakan adalah Regresi Linier Berganda dengan persamaan $Y = a + b_1X_1 + b_2X_2 + \dots + b_nX_n + e$, dipilih karena kesederhanaan, interpretasi koefisien yang mudah dipahami, dan efisiensi komputasi. Metode ini memungkinkan pengaruh setiap variabel terhadap harga tanah dijelaskan secara kuantitatif, yang penting bagi instansi pemerintah dan praktisi properti. Model dilatih menggunakan data latih dan diuji dengan data uji untuk mengevaluasi kinerja prediktifnya. Data kemudian dibagi menjadi 80% data latih dan 20% data uji menggunakan train-test split.

Tabel 1. Dataset Penelitian.

NO	Luas tanah (m2)	Lebar Jalan	Koordinat (Penentuan Jarak Ke Pusat Kota)		Luas Bangunan	Jumlah Lantai	Kelas Jalan	Harga Penawaran/ Transaksi (Rp.)
			X	Y				
1	374	10.0	345504	660519	275	1	Lokal	2.500.000.000
2	397	6.0	345495	660533	0	0	Kolektor	2.382.000.000
3	200	6.0	345917	660351	315	3	Kolektor	3.400.000.000
4	539	4.0	346056	660309	0	0	Lokal	2.100.000.000
5	584	4.0	345662	660138	275	1	Lokal	2.000.000.000

Evaluasi model dilakukan menggunakan tiga metrik standar, yaitu RMSE yang mengukur akar rata-rata kuadrat selisih prediksi dan aktual, MAE yang mengukur rata-rata selisih absolut dan lebih robust terhadap outlier, serta R^2 yang mengukur proporsi variasi variabel dependen yang dapat dijelaskan model. Kriteria model baik adalah $RMSE < \text{standar deviasi harga tanah}$ dan $R^2 > 0,7$, yang menandakan error prediksi masih dalam batas wajar dan model mampu menjelaskan sebagian besar variasi data.

Tahap akhir adalah *deployment* dengan menyimpan model dalam format .pkl menggunakan joblib dan membangun prototipe antarmuka berbasis Streamlit. Melalui antarmuka ini, pengguna memasukkan parameter tanah (luas tanah, lebar jalan, luas bangunan, jumlah lantai, kelas jalan, dan jarak dari pusat kota (m)) untuk mendapatkan estimasi harga.

Prototype ini memudahkan petugas instansi dan masyarakat dalam estimasi harga tanah secara cepat dan objektif.

4. HASIL DAN PEMBAHASAN

Hasil

Pada tahap ini, model Regresi Linier Berganda telah berhasil dibangun menggunakan total 966 data hasil pembersihan dari 1923 data kotor, dengan rincian pembagian 80% (772) data latih dan 20 % (194) data uji , dengan hasil evaluasi pada data uji menunjukkan kinerja yang baik. Model menghasilkan koefisien determinasi (R^2) sebesar 0,7767, yang berarti model mampu menjelaskan sekitar 77,67% variasi harga tanah berdasarkan variabel-variabel yang digunakan. Nilai RMSE yang diperoleh adalah Rp684.300.000, sedangkan MAE sebesar Rp503.900.000. Dengan standar deviasi harga tanah sebesar Rp1.493.000.000, nilai RMSE yang lebih kecil dari standar deviasi tersebut mengindikasikan bahwa model memiliki tingkat kesalahan prediksi yang masih dalam batas wajar

Masuk pada tahap pemodelan, variabel kategori kelas jalan diubah menggunakan *one-hot encoding* menjadi empat kolom biner: Kelas_Arteri, Kelas_Kolektor, Kelas_Lokal, dan Kelas_Setapak. Tabel 2 menyajikan data setelah encoding. Selain itu, fitur koordinat X dan Y digunakan untuk menghitung jarak *Euclidean* setiap bidang tanah terhadap titik pusat kota yang di sebut *centroid*. Fitur “jarak_ke_pusat_kota” ini menggantikan koordinat X dan Y karena lebih mudah diinterpretasikan. Selanjutnya, seluruh fitur numerik dinormalisasi menggunakan *StandardScaler* agar memiliki skala yang sama (rata-rata 0, standar deviasi 1), sehingga model regresi tidak bias terhadap variabel dengan skala lebih besar. Tabel 3 menyajikan data setelah normalisasi. Persamaan regresi yang dihasilkan adalah sebagai berikut :

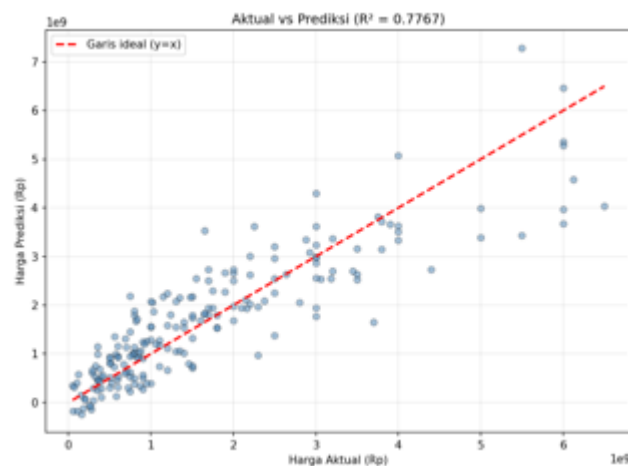
$$\text{Harga Tanah} = a + b_1(\text{Luas Tanah}) + b_2(\text{Lebar Jalan}) + b_3(\text{Jarak ke Pusat Kota}) + b_4(\text{Luas Bangunan}) + b_5(\text{Jumlah Lantai}) + b_6(\text{Kelas Jalan})$$

Tabel 2. Data Setelah Encoding.

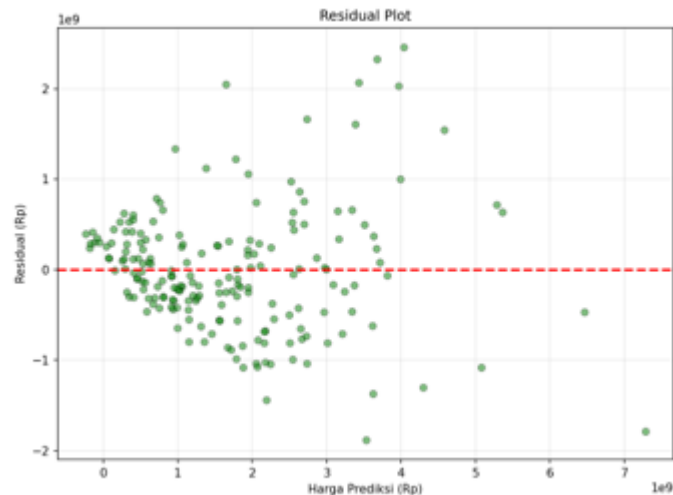
No	Luas Tanah	Lebar Jalan	Jarak_Ke_Pusat_Kota	Luas Bangunan	Jumlah Lantai	Kelas_Arteri	Kelas_Kolektor	Kelas_Lokal	Kelas_Stapak
1	347	10	3203.694118	200.0	1.0	True	False	False	False
2	106	6	3109.019942	106.0	1.0	False	False	True	False
3	210	6	3323.499361	170.0	1.0	False	True	False	False
4	104	4	3290.950167	95.0	1.0	False	False	True	False
5	200	4	3286.743221	300.0	2.0	False	False	True	False

Tabel 3. Data setelah normalisasi (*StandardScaler*).

Luas Tanah	Lebar Jalan	Jarak_Ke_Pusat_Kota	Luas Bangunan	Jumlah Lantai	Kelas_Arteri	Kelas_Kolektor	Kelas_Lokal	Kelas_Stapak
0.928370	1.124385	0.658950	0.206548	0.565539	3.577709	0.595683	1.311917	0.185098
0.639289	0.002128	0.569767	-0.536606	0.565539	0.279508	0.595683	0.762243	0.185098
0.037211	0.002128	0.771807	-0.030629	0.565539	0.279508	1.678744	1.311917	0.185098
0.652299	0.565385	0.741146	-0.623571	0.565539	0.279508	0.595683	0.762243	0.185098
0.027837	0.565385	0.737183	0.997137	1.219790	0.279508	0.595683	0.762243	0.185098

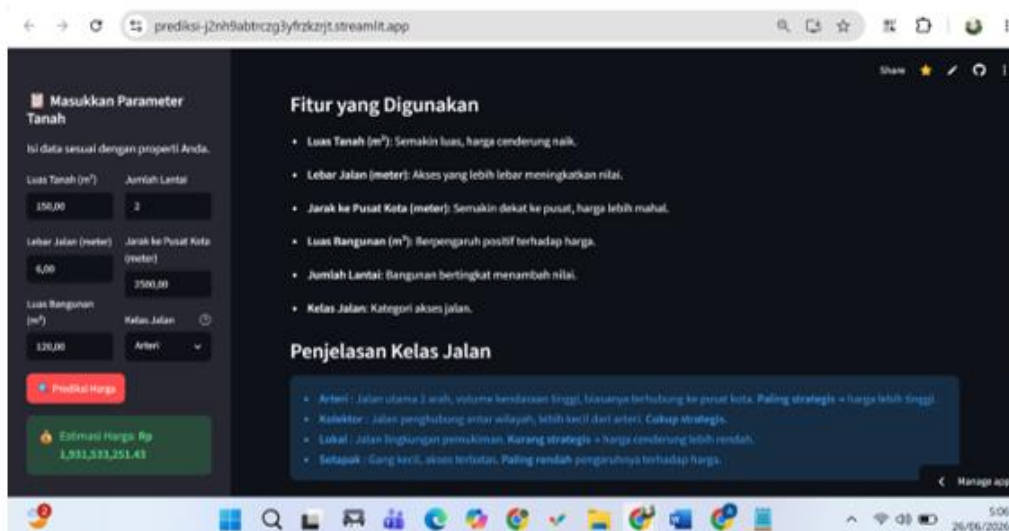
**Gambar 1.** Scatter Plot Antara Nilai Prediksi dan Nilai Actual.

Sumbu X menunjukkan harga aktual dan sumbu Y menunjukkan harga prediksi. Titik-titik berwarna biru merepresentasikan hasil prediksi model, sedangkan garis merah diagonal menunjukkan kondisi ideal di mana prediksi sama persis dengan aktual. Semakin dekat titik-titik dengan garis diagonal, semakin akurat model dalam melakukan prediksi. Berdasarkan gambar terlihat bahwa sebagian besar titik mengelilingi garis diagonal, yang mengindikasikan model memiliki akurasi yang cukup baik.



Gambar 2. *Residual Plot.*

Sumbu X menunjukkan nilai prediksi dan sumbu Y menunjukkan residual (selisih prediksi dan aktual). Titik-titik yang tersebar acak di sekitar garis nol tanpa membentuk pola tertentu mengindikasikan bahwa asumsi homoskedastisitas terpenuhi.



Gambar 3. Hasil *Deployment* menggunakan *Streamlit*.

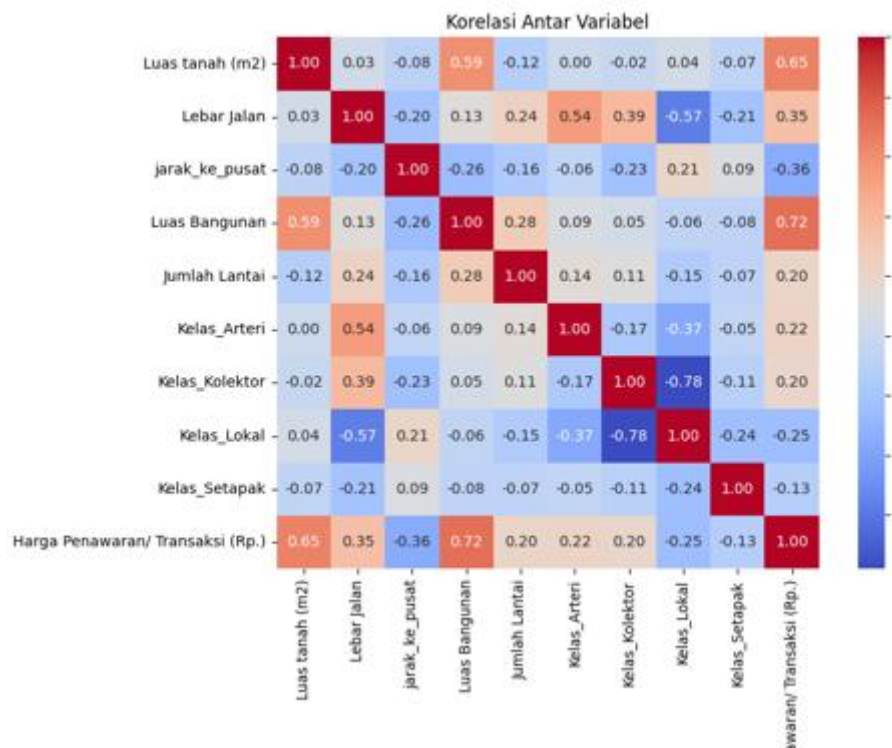
Model yang telah di uji diimplementasikan ke dalam prototipe berbasis *Streamlit*. Pengguna dapat memasukkan parameter luas tanah, lebar jalan, luas bangunan, jumlah lantai, kelas jalan, dan jarak ke pusat kota, lalu sistem menampilkan estimasi harga tanah secara *real-time*. *Prototipe* ini menjadi output dari hasil penelitian yang siap digunakan secara praktis.

Pembahasan

Proses penelitian diawali dengan pengumpulan data sebanyak 1.923 baris dari Kantor Pertanahan Kota Surakarta. Data tersebut melalui proses pembersihan yang meliputi penanganan nilai hilang menggunakan imputasi median dan penghapusan baris data yang memiliki nilai hilang pada variabel target, serta deteksi outlier menggunakan metode *Interquartile Range (IQR)*. Dari proses ini, diperoleh 966 data bersih yang siap digunakan.

Seleksi fitur dilakukan untuk memilih variabel yang paling kuat pengaruhnya terhadap harga tanah, menghasilkan enam variabel utama yaitu luas tanah, lebar jalan, luas bangunan, jumlah lantai, kelas jalan, dan jarak ke pusat kota. Data kemudian dibagi menjadi data latih sebesar 80% (772 data) dan data uji sebesar 20% (194 data).

Pada tahap pemodelan, variabel kategori kelas jalan diubah ke dalam bentuk numerik menggunakan *one-hot encoding* karena algoritma regresi linier hanya menerima input berupa angka. Metode ini menghasilkan empat kolom biner yaitu Kelas_Arteri, Kelas_Kolektor, Kelas_Lokal, dan Kelas_Setapak, seperti yang ditunjukkan pada Tabel 2. Selain itu, fitur koordinat X dan Y digunakan untuk menghitung jarak *Euclidean* setiap bidang tanah terhadap titik pusat kota, menghasilkan fitur jarak_ke_pusat_kota yang lebih mudah diinterpretasikan dalam bentuk satuan meter dan menggantikan koordinat X dan Y secara langsung. Seluruh fitur numerik kemudian dinormalisasi menggunakan StandardScaler agar memiliki skala yang sama (rata-rata 0, standar deviasi 1), sehingga model tidak bias terhadap variabel dengan skala yang lebih besar. Tabel 3 menyajikan contoh data setelah normalisasi.

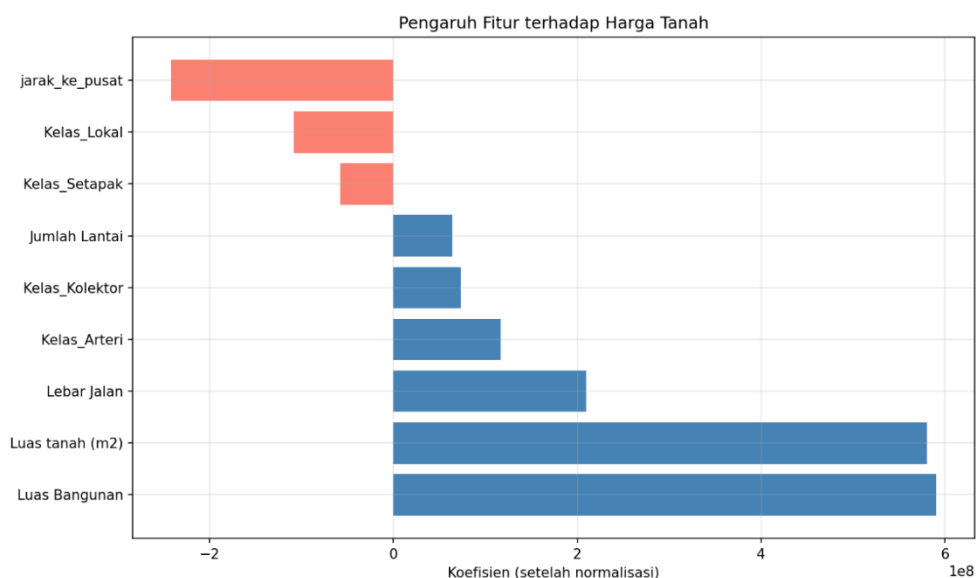


Gambar 4. Heatmap Korelasi Antar Variabel.

Untuk mempelajari hubungan antar variabel setelah proses *encoding* dan normalisasi, dilakukan analisis korelasi *Pearson* yang hasilnya disajikan pada Gambar 4. Variabel luas bangunan memiliki korelasi paling tinggi terhadap harga tanah (0,72), diikuti luas tanah (0,65), yang berarti semakin besar luas bangunan dan tanah, semakin tinggi harga jualnya.

Sebaliknya, jarak ke pusat kota berkorelasi negatif (-0,36), menunjukkan bahwa lokasi yang lebih jauh akan menurunkan harga tanah. Kelas jalan arteri (0,22) dan kolektor (0,20) berkorelasi positif, sedangkan lokal (-0,25) dan setapak (-0,13) berkorelasi negatif. Hal ini mengonfirmasi bahwa akses jalan yang lebih baik meningkatkan nilai tanah..

Model Regresi Linier Berganda dibangun menggunakan pustaka *Scikit-learn* dengan metode *Ordinary Least Squares* (OLS) yang meminimalkan jumlah kuadrat residual. Model dilatih menggunakan data latih sebanyak 772 data, kemudian diuji menggunakan data uji sebanyak 194 data. Berdasarkan hasil pengujian, model menghasilkan R^2 sebesar 0,7767, yang berarti model mampu menjelaskan sekitar 77,67% variasi harga tanah. Nilai RMSE yang diperoleh adalah Rp684.300.000, sedangkan MAE sebesar Rp503.900.000. Dengan standar deviasi harga tanah sebesar Rp1.493.000.000, nilai RMSE yang lebih kecil dari standar deviasi tersebut mengindikasikan bahwa model memiliki tingkat kesalahan prediksi yang masih dalam ambang batas wajar, sehingga model dinyatakan baik dan layak digunakan, meskipun dikemudian hari perlu adanya peningkatan dan inprovisasi kembali.



Gambar 5. Koefisien Regresi Normalisasi.

Koefisien regresi yang dihasilkan dinormalisasi untuk memudahkan perbandingan antar variabel yang memiliki satuan berbeda. Gambar 5 menunjukkan bahwa luas tanah dan luas bangunan memiliki pengaruh positif terbesar terhadap harga tanah (koefisien 5,5), diikuti lebar jalan (2,0), kelas kolektor (0,8), kelas arteri (0,7), dan jumlah lantai (0,5). Sebaliknya, jarak ke pusat kota memiliki pengaruh negatif terbesar (-2,8), diikuti kelas lokal (-1,0) dan kelas setapak (-0,5). Hal ini mengonfirmasi bahwa lokasi yang semakin jauh dari pusat kota dan akses jalan yang kurang baik cenderung menurunkan nilai tanah.

Untuk menguji kinerja model secara visual, dilakukan beberapa plot diagnostik. Scatter plot antara nilai prediksi dan nilai aktual pada data uji menunjukkan bahwa sebagian besar titik mengelilingi garis diagonal, yang mengindikasikan model memiliki akurasi yang cukup baik. Residual plot menunjukkan titik-titik yang tersebar acak di sekitar garis nol tanpa membentuk pola tertentu, yang mengindikasikan bahwa asumsi homoskedastisitas terpenuhi. Histogram residual menunjukkan distribusi yang mendekati kurva normal, sehingga asumsi normalitas residual juga terpenuhi. Dengan terpenuhinya asumsi-asumsi klasik regresi, model yang dihasilkan dapat diandalkan untuk melakukan prediksi.

Pengujian dengan data baru menunjukkan model mampu memberikan estimasi harga tanah yang mendekati nilai aktual dengan tingkat kesalahan rata-rata sekitar Rp503,9 juta (MAE) dan Rp684,3 juta (RMSE). Hal ini membuktikan model memiliki kemampuan generalisasi yang baik terhadap data baru yang belum pernah dilihat sebelumnya.

Kebaharuan yang di temukan dari penelitian ini terletak pada beberapa aspek. Pertama, penelitian ini menggunakan data transaksi dan penawaran tanah spesifik di Kota Surakarta tahun 2021–2024 yang diperoleh langsung dari Kantor Pertanahan Kota Surakarta, yang belum banyak dieksplorasi dalam penelitian prediksi harga tanah di Indonesia. Kedua, integrasi variabel spasial berupa koordinat X dan Y yang dikonversi menjadi jarak ke pusat kota memberikan dimensi geografis dalam prediksi yang jarang ditemukan pada penelitian sejenis. Ketiga, penerapan kerangka CRISP-DM secara sistematis dari pengolahan data hingga *deployment* menjadikan penelitian ini terdokumentasi dengan baik dan dapat direplikasi. Keempat, implementasi *prototipe* berbasis *Streamlit* memberikan nilai tambah praktis yang memudahkan pengguna, baik dari kalangan instansi pemerintah (BPN) maupun masyarakat umum, dalam melakukan estimasi harga tanah secara cepat dan objektif. Penelitian ini juga sejalan dengan Peraturan Menteri ATR/BPN Nomor 3 Tahun 2026 Pasal 29 yang mengizinkan penggunaan machine learning dalam penilaian tanah, sehingga penelitian ini dapat menjadi salah satu referensi dalam transformasi digital penilaian tanah di Indonesia.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa algoritma Regresi Linier Berganda berhasil diimplementasikan untuk memprediksi harga tanah di Kota Surakarta berdasarkan karakteristik fisik dan lingkungan. Model yang dibangun menggunakan 966 data bersih dari total 1.923 data awal dengan pembagian 80% data latih dan 20% data uji menghasilkan nilai R^2 sebesar 0,7767, yang berarti model mampu menjelaskan sekitar 77,67% variasi harga tanah.

Nilai RMSE sebesar Rp684.300.000 dan MAE sebesar Rp503.900.000, dengan standar deviasi harga tanah Rp1.493.000.000, menunjukkan bahwa model masih dalam batas wajar dan dinyatakan baik. Variabel luas tanah dan luas bangunan memiliki pengaruh positif terbesar terhadap harga tanah, sementara jarak ke pusat kota memiliki pengaruh negatif yang signifikan. Penelitian ini juga berhasil menerapkan kerangka CRISP-DM secara sistematis, menghasilkan model yang objektif, cepat, dan mudah diinterpretasikan, serta diimplementasikan dalam *prototipe* berbasis Streamlit yang memungkinkan pengguna memperoleh estimasi harga tanah secara real-time.

Akan tetapi penelitian ini memiliki beberapa keterbatasan yang perlu diperhatikan. Pertama, variabel yang digunakan dalam model masih terbatas pada karakteristik fisik dan spasial, sehingga belum mencakup faktor ekonomi dan sosial yang juga mempengaruhi harga tanah. Kedua, data yang digunakan hanya mencakup wilayah Kota Surakarta, sehingga hasil prediksi belum tentu berlaku umum untuk wilayah lain dengan karakteristik yang berbeda. Ketiga, *prototipe* yang dikembangkan masih dalam tahap demonstrasi dan belum terintegrasi dengan sistem yang ada di Kantor Pertanahan.

Untuk penelitian selanjutnya, disarankan untuk menambahkan variabel-variabel lain seperti kepadatan penduduk, tingkat ekonomi wilayah, dan kedekatan dengan fasilitas umum lainnya guna meningkatkan akurasi prediksi. Selain itu, penggunaan algoritma lain seperti *Random Forest* atau *Gradient Boosting* dapat dibandingkan untuk mengetahui algoritmana yang memberikan performa terbaik pada data harga tanah. Pengembangan lebih lanjut dari *prototipe* menjadi aplikasi yang terintegrasi dengan sistem Kantor Pertanahan juga sangat direkomendasikan agar dapat digunakan secara luas oleh petugas dan masyarakat. Terakhir, penelitian serupa dapat dilakukan di wilayah lain untuk menguji generalisasi model dan membandingkan faktor-faktor penentu harga tanah antar daerah.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Kantor Pertanahan Kota Surakarta yang telah memberikan akses data penelitian, serta kepada Ibu Aprilisa Arum Sari, S.T., M.Kom dan Bapak Pramono, S.Kom, M.Kom selaku dosen pembimbing yang telah memberikan bimbingan dan arahan selama proses penelitian. Ucapan terima kasih juga disampaikan kepada dewan penguji, Program Studi Teknik Informatika Universitas Duta Bangsa Surakarta, serta keluarga dan rekan-rekan atas dukungan yang telah diberikan. Penelitian ini merupakan bagian dari skripsi untuk memenuhi persyaratan menyelesaikan Program Strata 1 Teknik Informatika Universitas Duta Bangsa Surakarta.

DAFTAR REFERENSI

- Akbar Marhananda, A., & Ezra Fabiandika, F. (2026). SISTEM PREDIKSI HARGA RUMAH DI KOTA MALANG MENGGUNAKAN METODE RANDOM FOREST REGRESSOR BERBASIS MACHINE LEARNING. In *Jurnal Mahasiswa Teknik Informatika* (Vol. 10, Number 1).
- Alim. (2024). *ANALISIS DAMPAK PERUBAHAN NILAI TANAH PASCA PENGADAAN TANAH JALAN TOL YOGYAKARTA-BAWEN PROVINSI DAERAH ISTIMEWA YOGYAKARTA*.
- Alnuaimi, A. F. A. H., & Albaldawi, T. H. K. (2024). An overview of machine learning classification techniques. *BIO Web of Conferences*, 97. <https://doi.org/10.1051/bioconf/20249700133>
- 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>
- Hayya', M., Hana, F. M., & Prihandono, A. (2025). Perbandingan Metode K-Nearest Neighbor (KNN) dan Random Forest dalam Memprediksi Harga Rumah. *Sainteks*, 22(1), 99–108. <https://doi.org/10.30595/sainteks.v22i1.25998>
- Khoiriyah, S., & Fatah, Z. (2024). Penerapan Algoritma Linear Regression dalam Memprediksi Harga Rumah Menggunakan RapidMiner. *JUSIFOR: Jurnal Sistem Informasi Dan Informatika*, 3(2), 107–115. <https://doi.org/10.70609/jusifor.v3i2.5684>
- Lathifah, U., & Danar Dana, R. (2024). IMPLEMENTASI METODE LINEAR REGRESSION UNTUK PREDIKSI HARGA PROPERTI REAL ESTATE MENGGUNAKAN RAPIDMINER. In *Jurnal Mahasiswa Teknik Informatika* (Vol. 8, Number 1). <https://www.kaggle.com/datasets/oddyvirgantara/har>
- Nurani, A. T., Setiawan, A., & Susanto, B. (2023). Perbandingan Kinerja Regresi Decision Tree dan Regresi Linear Berganda untuk Prediksi BMI pada Dataset Asthma. *Jurnal Sains Dan Edukasi Sains*, 6(1), 34–43. <https://doi.org/10.24246/juses.v6i1p34-43>
- Pakusadewo, G., Yulianandha Maburur, A., Sai, S. S., & Susanti, D. B. (2023). Kajian Teknis Pembuatan Peta Zona Nilai Tanah Tahun 2021 Berdasarkan Standar Operasional Prosedur Kementerian Atr/Badan Pertanahan Nasional. *Prosiding SEMSINA*, 4(01), 200–205. <https://doi.org/10.36040/semsina.v4i01.8042>
- Parhusip, J., Julian, A. S., Hidayat, F. N., Souk, J. T., & Fakhri, N. (2025). *Pengaplikasian Algoritma Simple Linear Regression untuk Prediksi Harga Rumah di Jabodetabek Berdasarkan Fitur Lokasi dan Luas Bangunan*. 18(2), 110–117. <http://journal.stekom.ac.id/index.php/pixel>
- Putri, E. R., & Arianto, D. B. (2024). *Jurnal Nasional Teknologi dan Sistem Informasi Attribution-ShareAlike 4.0 International Some rights reserved Perbandingan Performa Algoritma Metode Bagging dan Boosting pada Prediksi Konsentrasi PM 10 di Jakarta Utara*. 01, 74–81.

- Risang Narindra, P., Fanani, K., Daffa Dzulfizar, F., Setiawan, A., Brilliant Nugraha, R., Wilayah dan Kota, P., & Teknik, F. (2024). *Pengaruh Sarana dan Prasarana terhadap Penentuan Zona Nilai Tanah (ZNT) berdasarkan Harga Pasar di Kecamatan Lowokwaru, Kota Malang*. 2(2), 67–82. <https://doi.org/10.59810/archimane/v2i2.44>
- Roustaei, N. (2024). Application and interpretation of linear-regression analysis. *Medical Hypothesis, Discovery, and Innovation in Ophthalmology*, 13(3), 151–159. <https://doi.org/10.51329/mehdiophthal1506>
- Salman, H. A., Kalakech, A., & Steiti, A. (2024). Random Forest Algorithm Overview. In *Babylonian Journal of Machine Learning* (Vol. 2024, pp. 69–79). Mesopotamian Academic Press. <https://doi.org/10.58496/BJML/2024/007>
- Sari, N. N., Anisah, T. T., & Fitriani, R. (2024). Implementasi Machine Learning Untuk Prediksi Harga Laptop Menggunakan Algoritma Regresi Linear Berganda. *Jurnal Manajemen Informatika (JAMIKA)*, 14(2), 162–177. <https://doi.org/10.34010/jamika.v14i2.12923>
- Septriaznu, 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>
- Singgalen, Y. A. (2023). Penerapan Metode CRISP-DM dalam Klasifikasi Data Ulasan Pengunjung Destinasi Danau Toba Menggunakan Algoritma Naïve Bayes Classifier (NBC) dan Decision Tree (DT). *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 7(3), 1551. <https://doi.org/10.30865/mib.v7i3.6461>
- Syafarina, G. A. (2023). *Implementasi Framework Streamlit Sebagai Prediksi Harga Jual Rumah Dengan Linear Regresi lain menjelaskan bahwa Pegujian akurasi dilakukan dengan mengukur Root Mean Squared Error (RMSE). Nilai rata-rata jual rumah yang akurat di Kota Banjarmasin , mengg.* 121–125. <https://doi.org/10.47002/metik.v7i2.680>
- Yulianto, C. (n.d.). *MODEL PENILAIAN TANAH MASSAL BERBASIS BIDANG TANAH MENGGUNAKAN ALGORITMA RANDOM FOREST DI KOTA SURAKARTA PARCEL-BASED MASS VALUATION USING RANDOM FOREST IN SURAKARTA CITY.*