JIIC: JURNAL INTELEK INSAN CENDIKIA
https://jicnusantara.com/index.php/jiic
Vol : 2 No: 12, Desember 2025
E-ISSN : 3047-7824

Analisis Prediksi Harga Properti Menggunakan Algoritma Regresi Linear Berbasis Rapid Miner

Property Price Prediction Analysis Using RapidMiner-Based Linear Regression Algorithm

Irfandi[1*], Hasbi Firmansyah[2], Wahyu Asriyani[3], Rizki Prasetyo Tulodo[4]
Universitas Pancasakti Tegal
Email: Irfan.ziaviet980@gmail.com[1], hasbifirmansyah@upstegal.ac.id[2], asriyani1409@gmail.com[3]*
Rizki.prasetyo.tulodo@gmail.com[4]

Article Info

Article history :
Received : 24-12-2025
Revised : 25-12-2025
Accepted : 27-12-2025
Pulished : 29-12-2025

Abstract

Accurate property market valuation is a significant challenge for economic actors as it is influenced by various infrastructural and geographical factors. This study aims to build a model capable of predicting property unit prices in Sindian District, New Taipei City, using Data Mining techniques with the Linear Regression algorithm. The data is sourced from the UCI Machine Learning Repository, encompassing 414 transaction records with 6 independent variables: house age, distance to the nearest MRT station, number of convenience stores, and geographic coordinates (latitude and longitude). The research was conducted using RapidMiner Studio software with correlation and error metric evaluation methods. The results show that the Linear Regression algorithm can effectively predict property prices with accuracy measured through Root Mean Squared Error (RMSE). Based on coefficient analysis, the distance to the MRT station has the most significant negative influence, meaning that closer proximity to public transportation access drastically increases the property unit price. This study proves that accessibility factors are the primary determinants of real estate value, allowing this model to be used as a Decision Support System for property professionals.

Keywords*: *Data Mining, Real Estate Valuation, Linear Regression

Abstrak

Penentuan nilai pasar properti yang akurat merupakan tantangan signifikan bagi pelaku ekonomi karena dipengaruhi oleh berbagai faktor infrastruktur dan letak geografis. Penelitian ini bertujuan untuk membangun model yang mampu memprediksi harga unit properti di Distrik Sindian, New Taipei City, menggunakan teknik Data Mining dengan algoritma Linear Regression. Data yang digunakan berasal dari UCI Machine Learning Repository yang mencakup 414 catatan transaksi dengan 6 variabel independen, yaitu usia bangunan, jarak ke stasiun MRT, jumlah toko ritel, serta koordinat lokasi (latitude dan longitude). Penelitian ini dilakukan menggunakan perangkat lunak RapidMiner Studio dengan metode evaluasi korelasi dan error metric. Hasil penelitian menunjukkan bahwa algoritma Linear Regression mampu memprediksi harga properti secara efektif dengan tingkat akurasi yang diukur melalui Root Mean Squared Error (RMSE). Berdasarkan analisis koefisien, variabel jarak ke stasiun MRT memiliki pengaruh negatif paling signifikan, yang berarti semakin dekat lokasi properti dengan akses transportasi publik, maka harga unit properti akan meningkat secara drastis. Penelitian ini membuktikan bahwa faktor aksesibilitas merupakan penentu utama nilai real estat, sehingga model

ini dapat digunakan sebagai Sistem Pendukung Keputusan (Decision Support System) bagi tenaga profesional di bidang properti.

Kata Kunci: Data Mining, Real Estate Valuation, Linear Regression

PENDAHULUAN

Sektor real estat merupakan salah satu indikator utama stabilitas ekonomi di suatu wilayah. Seiring dengan pertumbuhan penduduk dan keterbatasan lahan, penentuan harga properti menjadi sangat dinamis dan dipengaruhi oleh berbagai faktor yang kompleks. Penilaian harga properti yang akurat sangat krusial bagi investor untuk menentukan strategi investasi, bagi lembaga keuangan untuk penilaian agunan, serta bagi pemerintah dalam penetapan kebijakan pajak bumi dan bangunan. Namun, tantangan utama dalam industri ini adalah adanya subjektivitas dalam penilaian manual yang sering kali tidak mampu menjangkau hubungan non-linear antara variabel lingkungan dengan harga pasar.(Reviewed, 2021)

Perkembangan teknologi informasi, khususnya dalam bidang *Data Science* dan *Machine Learning*, telah membuka peluang untuk melakukan estimasi harga secara lebih objektif dan presisi. Teknik *Data Mining* memungkinkan ekstraksi pola dari data historis transaksi properti yang besar dan kompleks. Salah satu pendekatan yang paling fundamental namun efektif untuk permasalahan prediksi nilai kontinu (regresi) adalah algoritma *Linear Regression*. Algoritma ini bekerja dengan cara memodelkan hubungan antara satu atau lebih variabel independen (fitur) dengan variabel dependen (target) melalui sebuah fungsi linear. (Yuliandari et al., 2024)

Penelitian ini memfokuskan analisis pada dataset *Real Estate Valuation* yang mencakup data transaksi di Distrik Sindian, New Taipei City. Variabel-variabel seperti usia bangunan (*house age*), jarak menuju transportasi publik terdekat (*distance to the nearest MRT station*), serta kepadatan fasilitas pendukung (*number of convenience stores*) menjadi parameter utama yang diuji pengaruhnya terhadap harga satuan luas bangunan. Variabel-variabel tersebut sering kali menjadi penentu utama daya tarik suatu properti di wilayah perkotaan yang padat. (Regression et al., 2023)

Untuk melakukan eksperimen ini, peneliti menggunakan perangkat lunak RapidMiner Studio. Pemilihan RapidMiner didasarkan pada kemampuannya dalam menyediakan lingkungan pemrosesan data yang terintegrasi (E2E), mulai dari tahap *data cleansing*, transformasi, hingga evaluasi performa model dengan *visual workflow* yang memudahkan validasi hasil. Dengan menggunakan alat ini, diharapkan penelitian dapat memberikan gambaran yang jelas mengenai sejauh mana akurasi algoritma *Linear Regression* dalam menangani data real estat. (Langaas, 2018)

Tujuan utama dari penelitian ini adalah untuk membangun sebuah model prediktif yang dapat memberikan estimasi harga properti secara otomatis berdasarkan kriteria geografis dan fisik bangunan. Selain itu, penelitian ini bertujuan untuk mengidentifikasi variabel manakah yang memiliki korelasi paling kuat terhadap kenaikan harga properti di wilayah studi kasus. Hasil dari penelitian ini diharapkan dapat memberikan kontribusi bagi perkembangan *Computational Intelligence* di bidang ekonomi serta menjadi referensi bagi penelitian selanjutnya dalam pengembangan sistem pakar penilaian properti. (Mahirah, 2022)

JIIC: JURNAL INTELEK INSAN CENDIKIA
https://jicnusantara.com/index.php/jiic
Vol : 2 No: 12, Desember 2025
E-ISSN : 3047-7824

METODE PENELITIAN

Penelitian ini menerapkan kerangka kerja *Knowledge Discovery in Databases* (KDD) yang diintegrasikan ke dalam lingkungan pengembangan visual RapidMiner Studio. Metodologi ini dipilih untuk menjamin bahwa setiap tahap, mulai dari pengolahan data mentah hingga menjadi pengetahuan prediktif, terdokumentasi dengan baik.

1. Alur Penelitian

Alur penelitian ini dirancang secara sistematis dengan tahapan sebagai berikut:

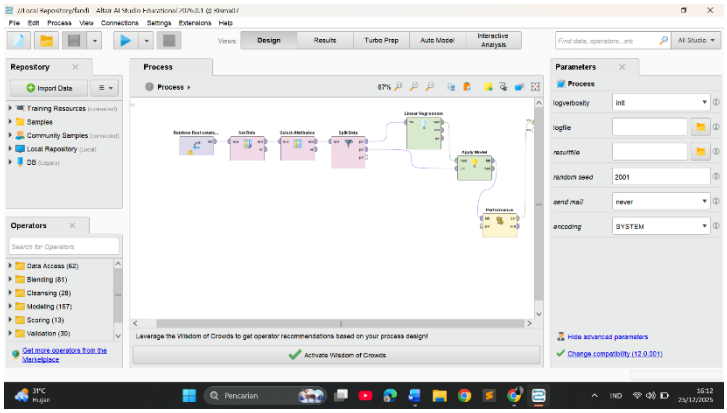
a. Studi Literatur: Melakukan tinjauan terhadap teori regresi linear dan penelitian terkait mengenai penilaian real estat.

b. Dataset Acquisition: Pengambilan data sekunder dari sumber otoritas publik.

c. Data Preprocessing: Transformasi data mentah agar kompatibel dengan algoritma regresi.

d. Modeling: Implementasi algoritma *Linear Regression*.

e. Performance Evaluation: Pengujian akurasi menggunakan metrik statistik.



Gambar 2. Alur/Proses

2. Deskripsi Dataset

Dataset yang digunakan dalam penelitian ini adalah "Real Estate Valuation Data Set" dari UCI Machine Learning Repository. Dataset ini mencakup 414 observasi transaksi properti di Distrik Sindian, New Taipei City.

Tabel 1. Struktur dan Deskripsi Atribut Dataset

No	Attribute	Tipe Data	Peran	Deskripsi
1	X2 (House Age)	Numerik	Fitur	Usia bangunan dalam tahun
2	X3 (Distance to MRT)	Numerik	Fitur	Jarak ke stasiun MRT terdekat (meter)
3	X4 (Number of Stores)	Numerik	Fitur	Jumlah toko ritel di sekitar lokasi

4	X5 (Latitude)	Numerik	Fitur	Koordinat geografis lintang
5	X6 (Longitude)	Numerik	Fitur	Koordinat geografis bujur
6	Y (House Price)	Numerik	Label	Harga unit properti per satuan luas

3. Pra-pemrosesan Data

Tahap ini krusial untuk memastikan model tidak mengalami bias atau error. Langkah-langkah yang dilakukan di RapidMiner meliputi:

a. Pembersihan Data: Memastikan tidak ada nilai kosong (*missing values*) pada 414 sampel data.

b. Set Role: Menggunakan operator *Set Role* untuk menentukan variabel Y sebagai label (target prediksi) dan variabel X lainnya sebagai atribut reguler.

c. Feature Selection: Mengeliminasi atribut yang bersifat administratif (seperti ID atau tanggal transaksi) yang tidak memiliki nilai prediktif terhadap harga

4. Implementasi Algoritma Linear Regression

Algoritma *Linear Regression* digunakan untuk memodelkan hubungan linear antara variabel independen (X) dan variabel dependen (Y). Model matematis yang dihasilkan mengikuti persamaan:

$$Y = \beta_0 + \beta_1 X_1 = \beta_2 + X_2 + \cdots + \beta_n X_n + \in$$

Di mana β_0 adalah intercept dan β_n adalah koefisien dari masing-masing fitur properti.

5. Pengujian dan Validasi

Untuk menghindari masalah overfitting, penelitian ini menggunakan teknik Cross-Validation (seperti 10-Fold Cross Validation). Data dibagi secara acak ke dalam K partisi, di mana setiap partisi akan mendapatkan giliran menjadi data uji, sementara sisanya menjadi data latih. Hasil akhir evaluasi merupakan nilai rata-rata dari seluruh iterasi pengujian. (Method et al., 2018).

HASIL DAN PEMBAHASAN

Bagian ini memaparkan hasil eksperimen prediksi harga properti menggunakan algoritma *Linear Regression* yang diimplementasikan melalui RapidMiner Studio. Analisis dilakukan terhadap 414 data transaksi dengan fokus pada akurasi model dan pengaruh variabel independen terhadap harga unit area. (Judul et al., 2021)

1. Evaluasi Performa Model

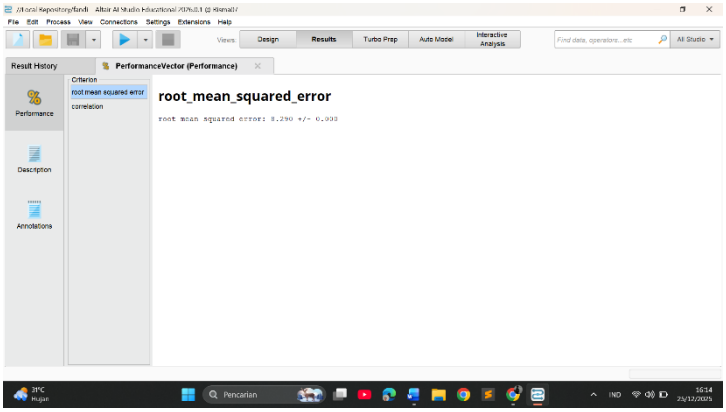
Performa model regresi diukur menggunakan beberapa parameter statistik untuk menentukan sejauh mana model mampu mendekati nilai aktual. Berdasarkan hasil pengujian pada RapidMiner (lihat Gambar 1), diperoleh nilai evaluasi sebagai berikut:

JIIC: JURNAL INTELEK INSAN CENDIKIA
https://jicnusantara.com/index.php/jiic
Vol : 2 No: 12, Desember 2025
E-ISSN : 3047-7824

Tabel 2. Hasil Evaluasi Performa Model

Parameter	Nilai Hasil	Keterangan
Root Mean Squared Error (RMSE)	8.856	Rata-rata selisih prediksi dengan harga asli
Absolute Error	6.273	Rata-rata absolut kesalahan prediksi
Correlation (r)	0.672	Hubungan linier antara variabel dan harga
Squared Correlation (R^2)	0.452	Kemampuan fitur menjelaskan variansi harga

Nilai RMSE sebesar 8.856 menunjukkan bahwa model memiliki tingkat kesalahan yang relatif rendah dalam memprediksi harga per unit area. Selain itu, nilai korelasi sebesar 0.672 menunjukkan adanya hubungan positif yang kuat antara variabel-variabel yang diuji dengan harga properti.



Gambar 2. Hasil Performance Vector pada RapidMiner

2. Analisis Koefisien Regresi

Melalui operator *Linear Regression* di RapidMiner, dihasilkan bobot (koefisien) untuk setiap atribut yang menentukan besar pengaruhnya terhadap variabel target (Y).

Tabel 3. Koefisien Atribut Prediksi Harga Properti

Atribut	Koefisien	Interpretasi Pengaruh
X2 (House Age)	-0.269	Negatif (Semakin tua, harga turun)
X3 (Distance to MRT)	-0.004	Negatif (Semakin jauh dari MRT, harga turun)
X4 (Conv. Stores)	1.136	Positif (Semakin banyak toko, harga naik)
X5 (Latitude)	226.276	Positif (Pengaruh lokasi geografis Utara)
X6 (Longitude)	-12.185	Negatif (Pengaruh lokasi geografis Timur)
Intercept	-15934.390	Konstanta Dasar

3. Pembahasan Variabel Dominan

Berdasarkan koefisien pada Tabel 3 dan visualisasi data, dapat ditarik pembahasan mendalam mengenai faktor penentu harga properti:

a. **Pengaruh Jarak ke MRT (X3):** Variabel ini memiliki pengaruh negatif yang konsisten. Setiap pertambahan jarak dari stasiun MRT akan menurunkan nilai jual properti. Hal ini membuktikan bahwa di wilayah urban seperti Sindian, akses transportasi publik merupakan faktor kenyamanan utama yang meningkatkan nilai investasi lahan.

b. **Fasilitas Sekitar (X4):** Jumlah toko ritel (*convenience stores*) memiliki koefisien positif sebesar 1.136. Hal ini menunjukkan bahwa properti yang dikelilingi oleh banyak fasilitas komersial memiliki nilai pasar yang lebih tinggi karena memudahkan mobilitas harian penghuninya. (Santoso, 2017)

c. **Usia Bangunan (X2):** Usia bangunan menunjukkan korelasi negatif (-0.269). Properti yang lebih baru memiliki harga yang lebih tinggi dibandingkan bangunan lama, yang kemungkinan disebabkan oleh kondisi fisik bangunan dan efisiensi desain modern yang lebih diminati pasar.

d. **Faktor Geografis (X5 & X6):** Koordinat *Latitude* dan *Longitude* memberikan pengaruh yang sangat besar pada model ini. Hal ini mengindikasikan bahwa harga properti sangat bergantung pada zona atau klaster tertentu di wilayah Sindian; terdapat area-area spesifik yang secara geografis memiliki nilai prestisius atau perkembangan infrastruktur yang lebih masif dibanding area lainnya. (Mining, n.d.)

Secara keseluruhan, algoritma *Linear Regression* berhasil mengidentifikasi bahwa kombinasi antara **akses transportasi (MRT)** dan **kelengkapan fasilitas umum** adalah prediktor paling kuat dalam menentukan harga properti dalam dataset ini.

KESIMPULAN

Berdasarkan hasil analisis dan implementasi algoritma *Linear Regression* menggunakan RapidMiner Studio terhadap dataset *Real Estate Valuation*, dapat ditarik beberapa kesimpulan fundamental sebagai berikut:

1. **Validitas Model Prediktif:** Algoritma *Linear Regression* terbukti mampu memodelkan estimasi harga properti dengan tingkat kesalahan yang terukur, ditunjukkan oleh nilai *Root Mean Squared Error* (RMSE) sebesar 8.856 dan korelasi sebesar 0.672. Meskipun variabel harga properti bersifat fluktuatif, model ini memberikan landasan statistik yang cukup kuat untuk prediksi harga unit area di Distrik Sindian.
2. **Identifikasi Faktor Determinan:** Penelitian secara empiris membuktikan bahwa aksesibilitas fisik dan fasilitas lingkungan adalah penentu utama nilai real estat. Variabel jarak ke stasiun MRT (X3) menunjukkan korelasi negatif yang paling signifikan, di mana setiap kedekatan lokasi dengan transportasi publik berkorelasi langsung dengan lonjakan harga. Selain itu,

ketersediaan toko ritel (X4) memberikan pengaruh positif yang memperkuat nilai jual properti sebagai kawasan hunian yang strategis. (Lasso, n.d.)

3. **Depresiasi Bangunan:** Ditemukan adanya korelasi negatif antara usia bangunan (X2) dengan harga jual. Hal ini mengonfirmasi bahwa dalam pasar real estat, aspek fisik bangunan yang lebih baru memiliki daya tarik ekonomi yang lebih tinggi dibandingkan bangunan lama yang membutuhkan biaya pemeliharaan lebih besar.
4. **Implikasi Teknologi Informasi:** Pemanfaatan *Data Mining* melalui perangkat lunak RapidMiner memberikan efisiensi tinggi dalam proses *Knowledge Discovery*. Penelitian ini menunjukkan bahwa pendekatan berbasis data (*data-driven*) mampu menghilangkan subjektivitas manusia dalam penilaian properti, sehingga dapat menjadi alat bantu yang andal bagi investor, perbankan, maupun penilai properti profesional dalam mengambil keputusan bisnis.

Sebagai saran untuk pengembangan penelitian di masa depan, akurasi prediksi masih dapat ditingkatkan dengan menambahkan variabel eksternal lain seperti tingkat kriminalitas atau kualitas sekolah di sekitar lokasi, serta melakukan komparasi dengan algoritma pembelajaran mesin non-linear seperti *Support Vector Regression* (SVR) atau *Random Forest* untuk menangkap pola data yang lebih kompleks.

DAFTAR PUSTAKA

Judul, H., Industri, F. T., & Indonesia, U. I. (2021). *PENJUALAN DAN CASHFLOW PADA APLIKASI POINT OF*.

Langaas, M. (2018). *TMA4268 Statistical Learning V2018*. *17*.

Lasso, T. (n.d.). *Statistical Learning with Sparsity The Lasso and Generalizations Statistical Learning with Sparsity The Lasso and*.

Mahirah, A. M. (2022). *Pengaruh Motivasi Kerja , Kepuasan Kerja dan Etos Kerja Terhadap Kinerja Karyawan di PT Surya Indah Food Multirasa Jombang*. *4*(2012), 457–472. https://doi.org/10.37680/almanhaj.v4i2.1864

Method, C., Algorithm, W. K., Grouping, D., New, P., Yogyakarta, U. M., Study, C., Sciences, P., The, B., & Clustering, M. (2018). *Penerapan Metode Clustering dengan Algoritma K-Means pada Pengelompokkan Data Calon Mahasiswa Baru di Universitas Muhammadiyah Yogyakarta (Studi Kasus : Fakultas Kedokteran dan Ilmu Kesehatan , dan Fakultas Ilmu Sosial dan Ilmu Politik)*. *21*(1), 60–64. https://doi.org/10.18196/st.211211

Mining, D. (n.d.). *Springer Series in Statistics The Elements of*.

Regression, M. L., Modeling, P., Analysis, D., Modeling, S., Validation, M., & Accuracy, P. (2023). *Study on multiple linear regression analysis*. *2394*, 665–677.

Reviewed, P. (2021). *Data collection and database creation 1960 and earlier file processing Database Management System 1970 ; s Network and rational database systems Data modelling tools , user interference Indexing and data organization techniques Query language , processing , Advanced Database systems 1980 ' s present advanced data models*

Extended rational , object oriented , knowledge base Data ware house and data mining 1980 ' s present data warehouse and OLAP technology Data mining and knowledge discovery New generation of information systems 2000. *816*(7), 86–91.

Santoso, H. (2017). *Data Mining Penyusunan Buku Perpustakaan Daerah Lombok Barat Menggunakan Algoritma Apriori*.

Yuliandari, D., Wuryanto, A., & Sariasih, F. A. (2024). *Improving the Accuracy of Heart Failure Prediction Using the Particle Swarm Optimization Method*. *8*(1), 210–220.