**Jurnal Riset Informatika dan Teknologi Informasi (JRITI)** ISSN (Print) : 3024-8167
Jejaring Penelitian dan Pengabdian Masyarakat (JPPM)
Vol. 1, No. 3, April – Juli 2024, Hal. 111 - 115
Web : https://ejurnal.jejaringppm.org/index.php/jriti

# Penerapan Decision Tree Regression dalam Memprediksi Harga Rumah di Provinsi Jawa Barat

**Alif Izzudin Ramadhan[1], Nafis Almajid[2] dan Yanuar Ginting[3]**
Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Bhayangkara Jakarta Raya
Email : alifizudin7758@gmail.com, nafisalmajid19@gmail.com, yanuarginting@gmail.com

## ABSTRAKSI

Rumah memiliki peran yang penting sebagai kebutuhan pokok, tidak hanya sebagai tempat berlindung dan beristirahat. Selain itu, nilai jual rumah juga sangat dipengaruhi oleh faktor-faktor lingkungan seperti kedekatan dengan pusat perbelanjaan, perkantoran, luasnya tanah, dan lain sebagainya. Untuk mendapatkan pemahaman yang lebih akurat dan mampu memprediksi nilai jual rumah, penelitian ini mengusulkan penerapan metode *machine learning*. Penelitian ini melibatkan penggunaan tiga algoritma machine learning, yakni *multiple linear regression, decision tree regression, dan linear support vector regression*, untuk memprediksi harga rumah di Provinsi Jawa Barat. Dengan memanfaatkan data historis dan berbagai fitur terkait, seperti jumlah kamar, jumlah wc, jumlah garasi, luas tanah, dan lebar tanah. Algoritma *machine learning* akan melakukan analisis yang kompleks dan memberikan estimasi harga rumah yang lebih akurat. Diharapkan hasil penelitian ini dapat memberikan wawasan berharga bagi para pemangku kepentingan dalam industri properti di Provinsi Jawa Barat. Pendekatan machine learning diharapkan dapat meningkatkan kemampuan dalam memprediksi harga rumah dan memberikan informasi yang bermanfaat bagi pembeli, penjual, serta pihak terkait lainnya dalam mengambil keputusan yang lebih baik dalam transaksi properti.

***Kata Kunci:*** *Harga Rumah, Prediksi, Multiple Linear Regression, Decision Tree Regressor, Linear Support Vector Regressor, Jawa Barat*

## ABSTRACT

*A house plays an important role as a basic necessity, not only as a shelter and place of rest. Additionally, the selling price of a house is greatly influenced by environmental factors such as proximity to shopping centers, office buildings, land size, and others. To gain a more accurate understanding and predict the selling price of houses, this research proposes the application of machine learning methods. This study involves the use of three machine learning algorithms: multiple linear regression, decision tree regression, and linear support vector regression to predict house prices in West Java Province. By utilizing historical data and various relevant features, such as the number of rooms, bathrooms, garages, land area, and width, the machine learning algorithms will conduct complex analyses and provide more accurate price estimations. The expected outcome of this research is to provide valuable insights for stakeholders in the property industry in West Java Province. The adoption of machine learning approaches is anticipated to enhance the ability to predict house prices and provide useful information for buyers, sellers, and other relevant parties in making better decisions in property transactions.*

***Keywords:*** *House Price, Prediction, Multiple Linear Regression, Decision Tree Regressor, Linear Support Vector Regressor, West Java*

**Penulis Korespondensi**
*Alif Izzudin Ramadhan*

Tanggal Submit : 14/07/2023
Tanggal Diterima : 11/07/2024
Tanggal Terbit : 26/07/2024



## I. PENDAHULUAN

Rumah adalah kebutuhan pokok selain sandang dan pangan yang tidak dapat dihindari. Berdasarkan data dari Badan Pusat Statistik tahun 2019-2021, Provinsi Jawa Barat merupakan Provinsi kedua dengan tingkat kepadatan penduduk tertinggi di Indonesia dengan rata-rata populasi 1.379 jiwa/km², hal ini menyebabkan meningkatnya kebutuhan rumah di

Provinsi Jawa Barat [1] serta makin kompetitifnya harga perumahaan.

Harga rumah dapat naik dan turun, terlebih lagi dekat dengan pusat perbelanjaan, pusat perkantoran, pusat sarana transportasi, dan fasilitas-fasilitas didalamnya. Hal ini yang sering dijadikan alasan para developer untuk menetapkan harga sampai mencapai harga yang tidak masuk akal. Permasalahan yang dihadapi developer saat ini adalah bagaimana menentukan harga rumah yang sesuai oleh penjual properti rumah berdasarkan fasilitas-fasilitas yang tersedia namun tetap memberikan harga yang kompetitif [2].

Jika permasalahan dalam menentukan harga rumah yang sesuai dan kompetitif tidak dapat diselesaikan, maka hal ini dapat berdampak pada kenaikan harga rumah yang tidak masuk akal sehingga masyarakat mengalami kesulitan untuk membeli rumah sesuai kebutuhan dan kemampuan. Selain itu, ketidakmampuan developer dalam menentukan harga rumah yang sesuai dapat menghambat pertumbuhan industri properti dan dapat mempengaruhi stabilitas pasar properti secara keseluruhan. Ada banyak cara yang dapat digunakan untuk menentukan harga rumah dengan sesuai, terutama dengan pendekatan komputasi, salah satunya *decision tree regression* (DTR).

DTR adalah jenis pohon keputusan yang digunakan untuk tugas regresi yang dapat digunakan untuk memprediksi output dengan nilai kontinu daripada output dengan nilai diskrit. Algoritma ini biasanya digunakan untuk menyelesaikan permasalahan seperti prediksi harga mobil [3], prediksi harga real estate [4], prediksi harga tiket pesawat [5], prediksi radiasi matahari [6] prediksi harga cryptocurrency [7], dan prediksi harga saham Indonesia [8].

Berdasarkan masalah tersebut. Digunakan pendekatan *artificial intelligence* dengan metode *machine learning* dalam membuat model prediksi harga rumah menggunakan algoritma *multiple linear regression, decision tree regression, dan linear support vector regression*. Tujuan utama penelitian ini adalah untuk mempelajari dan mengidentifikasi bagaimana *machine learning* dapat digunakan dalam memprediksi harga rumah berdasarkan fitur yang tersedia. Adapun manfaat dari penelitian ini adalah untuk menambah pengetahuan dan pemahaman tentang penerapan machine learning dapat digunakan para developer dalam memprediksi harga rumah.

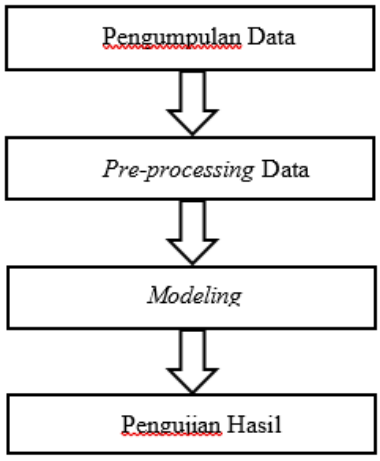
## II. PENELITIAN YANG TERKAIT

Pembahasan prediksi harga rumah menggunakan pendekatan *artificial intelligence* telah banyak digunakan oleh beberapa peneliti. Sebagian besar ilmuwan menggunakan *supervised learning* sebagai sarana untuk memprediksi harga rumah dan hal tersebut terbukti bekerja dengan baik dengan algoritma *linear regression* [9], d*ecision tree regression* [10], dan support vector regression [11].

Beberapa penelitian diluar topik prediksi harga rumah menggunakan algoritma *multiple linear regression, decision tree regression, dan support vector regression* misalnya untuk riset di bidang investasi, seperti prediksi pasar saham [12], prediksi harga saham gabungan (IHSG) [13], dan prediksi Harga Saham Indonesia pada masa Covid-19 [14]. Di bidang meteorologi dan klimatologi, seperti prediksi kekuatan gempa bumi [15], dan prediksi curah hujan berpotensi banjir [16]. Di bidang perkebunan dan pertanian, seperti prediksi harga kelapa sawit [17], prediksi harga kedelai lokal dan kedelai impor [18], dan prediksi harga cabai merah [19]. Walaupun banyak pendekatan lain guna memprediksi harga rumah seperti *random forest, neural network*, serta *backpropagation*. Akan tetapi pada penelitian ini algoritma *multiple linear regression, decision tree regression, serta linear support vector machine* regression digunakan sebagai komparasi dalam memprediksi harga rumah.

## III. METODE PENELITIAN

Dalam penelitian ini data yang digunakan adalah data kuantitatif atau data yang berupa bilangan, lalu pada tahap metode penelitian yang dilakukan adalah dengan pengumpulan data dengan variabel yang digunakan, *pre-processing* data atau tahap pengolahan awal data, lalu melakukan penerapan algoritma *machine learning multiple linear regression, decision tree regression*, dan *linear support vector machine regression* pada data, serta melakukan komparasi hasil dari pengujian algoritma yang digunakan.




Gambar. 1 Proses pengolahan data.

### *A. Pengumpulan Data*

Dalam penelitian ini menggunakan data yang dikumpulkan melalui *web scrapping* dari sebuah website jual beli rumah terpercaya. Dari hasil survei terhadap pengembang rumah, terdapat lima faktor yang memiliki pengaruh terhadap harga sebuah rumah, yaitu ukuran lahan, lebar bangunan, jumlah kamar tidur, jumlah kamar mandi, dan ketersediaan tempat parkir mobil [20]. Jumlah data yang terkumpul adalah sebanyak 3947 data.

*B. Pre-processing Data*

Langkah awal dalam pengolahan data ialah melakukan *pre-processing* guna memastikan jika metode yang digunakan bisa berjalan dengan maksimal [21], tahapan ini tidak bisa dilewatkan dalam pengolahan data. pengolahan data yang dilakukan antara lain *cleaning* data atau proses menghapus serta memperbaiki informasi yang rusak maupun informasi yang tidak relevan, *transforming* data atau proses mengubah skala pengukuran data awal menjadi format yang berbeda, sehingga data bisa memenuhi asumsi-asumsi yang digunakan dalam analisis ragam dengan memanfaatkan *minmax scaling*.

*Minmax scaling* merupakan metode transformasi data yang mencakup pengubahan data dari satu rentang ke rentang yang berbeda, sambil tetap mempertahankan korelasi data yang asli [22]. Umumnya metode ini merubah data menjadi skala jangkauan 0 s.d. 1 sehingga rumusnya menjadi (1):

$$x' = \frac{x - min(x)}{max(x) - min(x)} \quad (1)$$

*C. Modeling*

Pada tahap *modeling*, data yang sudah di *pre-processing* akan di proses menggunakan algoritma *decision tree regression* sehingga mendapatkan sebuah pola dari data tersebut. Data harga rumah yang diperoleh berkisar antara Rp200.000.000 s.d. Rp2.000.000.000. Data kemudian dibagi menjadi dua jenis, yang pertama terdiri dari 80% data latih dan 20% data uji. Data latih kemudian diimplementasikan ke dalam model *decision tree regression* untuk kebutuhan model regresi. Ketiga model ini akan dibandingkan untuk mengetahui kinerja mana yang lebih baik dan dapat dijadikan acuan untuk memprediksi harga rumah.

1) Decision Tree Regression

*Decision Tree* merupakan suatu metode statistik yang menggunakan struktur pohon untuk menghasilkan model prediksi berdasarkan *serangkaian* aturan dan keputusan yang diambil dari data latih. Algoritma ini mencari aturan terbaik yang dapat memisahkan data latih ke dalam kelompok-kelompok yang homogen atau serupa berdasarkan atribut-atribut yang ada. Adapun perhitungan *decision tree regression* yang digunakan dinyatakan pada (2).

*Mean Squared Error*:

$$\underline{y} = \frac{1}{n_i} \sum_{y \varepsilon Q_i} y$$

$$H(Q_i) = \frac{1}{n_i} \sum_{y \varepsilon Q_i} (y - \underline{y}_i)^2 \quad (2)$$

Di mana:

$y$ = Nilai aktual output
$\underline{y}$ = Nilai prediksi output
$n_i$ = Jumlah partisi data
$H(Q_i)$ = Mean squared error

*D. Pengujian Hasil*

Setelah dilakukan pengujian terhadap ketiga model algoritma yang telah disebutkan sebelumnya, didapati bahwa algoritma *decision tree* memperoleh nilai R2 Score yang paling tinggi yaitu sebesar 0.86, sedangkan algoritma linear regression dan support vector machine hanya memperoleh nilai R2 Score sebesar 0.78.

Tabel I. Hasil Pengujian Model

| Model | R2 Score | MAE | MSE | RMSE |
|---|---|---|---|---|
| Decision Tree | 0,8689 | 91.436.523 | 2.3421 | 153.042.015 |
| Linear Regression | 0,7796 | 148.727.723 | 3.9370 | 198.420.565 |
| Linear SVM | 0,7795 | 148.888.679 | 3.9384 | 198.454.492 |

Untuk metrik evaluasi MAE, nilai terendah ditemukan pada tabel I model *decision tree regression* dengan nilai sebesar 91.436.523. Model *multiple linear regression* memiliki nilai MAE sebesar 148.727.723, sedangkan model *linear support vector machine regression* memiliki nilai MAE sebesar 148.888.679.

Berdasarkan hasil evaluasi performa model, dapat dilihat pada tabel I bahwa nilai MAE terendah yang diperoleh adalah 91.436.523, sedangkan rata-rata harga rumah di dataset adalah 751.958.937. Untuk mengetahui apakah nilai MAE tersebut baik atau buruk, maka perlu dilakukan perhitungan persentase kesalahan prediksi dengan membagi nilai MAE dengan rata-rata harga rumah di dataset tersebut, yaitu:

$$91.436.523 \div 751.958.937 \times 100\% = 12.15\%$$

Dalam hal ini, persentase kesalahan prediksi sebesar 12.15% terletak di antara kisaran nilai MAE yang dianggap baik dalam konteks prediksi harga rumah pada dataset tersebut, yaitu antara 10% hingga 20% dari rata-rata harga rumah di dataset.

## IV. HASIL DAN PEMBAHASAN

Hasil penelitian menunjukkan bahwa decision tree adalah algoritma terbaik dalam memprediksi harga rumah dibandingkan dengan *multiple linear regression* dan *linear support vector machine regression*. Hal ini disebabkan oleh kemampuan *decision tree regression* dalam menemukan hubungan kompleks antara fitur-fitur pada dataset dan memahami interaksi antara fitur-fitur tersebut.

*Decision tree regression* memiliki kemampuan untuk memisahkan dataset menjadi sub-grup yang lebih kecil berdasarkan fitur-fiturnya yang paling penting. Kemampuan ini sangat berguna dalam kasus dataset yang kompleks dengan banyak fitur yang saling berkaitan. Dalam kasus ini, *Decision Tree*

*Regression* dapat menemukan pola-pola yang lebih kompleks dan mencapai performa yang lebih baik dibandingkan dengan *multiple linear regression* dan *linear support vector machine regression*.

*Multiple linear regression* memiliki asumsi-asumsi tertentu yang dibuat dalam modelnya, seperti linearitas yang dapat membuat performa model menurun jika asumsi-asumsi tersebut tidak terpenuhi. Selain itu, linear regression juga dapat mengalami masalah dalam menangani fitur-fitur yang saling berkaitan atau tidak memiliki pengaruh yang signifikan terhadap variabel target.

Sementara itu, *linear support vector machine regression* juga memiliki kemampuan dalam menangani dataset yang kompleks dan fitur-fitur yang berkaitan satu sama lain. Namun, *linear support vector machine regression* dapat mengalami masalah jika terdapat banyak fitur yang tidak relevan atau jika terdapat banyak outlier pada dataset.

## V. KESIMPULAN

Berdasarkan hasil pengujian pada dataset harga rumah dengan menggunakan tiga model algoritma yaitu *decision tree regression*, *multiple linear regression*, dan *linear support vector machine regression*, didapatkan kesimpulan bahwa model *decision tree regression* menghasilkan nilai R2 Score yang paling tinggi, yaitu sebesar 0.86, sementara model *multiple linear regression* dan *linear support vector machine regression* hanya memperoleh nilai R2 Score sebesar 0.78. Selain itu, untuk metrik evaluasi MAE, model *decision tree regression* juga memiliki nilai yang paling rendah yaitu sebesar 85.639.950, sementara model *multiple linear regression* dan *linear support vector machine regression* memiliki nilai MAE yang lebih tinggi, yaitu masing-masing sebesar 148.727.700 dan 148.888.700. Oleh karena itu, dapat disimpulkan bahwa model *decision tree regression* adalah model terbaik untuk digunakan dalam memprediksi harga rumah pada dataset yang telah digunakan.

## DAFTAR PUSTAKA


[1] Badan Pusat Statistik, "*Kepadatan Penduduk menurut Provinsi (jiwa/km2),"* Badan Pusat Statistik. [Online]. Available: https://www.bps.go.id/indicator/12/141/1/kepadatan-penduduk-menurut-provinsi.html. [Accessed: 14-Jul-2023].

[2] M. R. Fahlepi and A. Widjaja, "Penerapan Metode Multiple Linear Regression Untuk Prediksi Harga Sewa Kamar Kost," *Jurnal STRATEGI-Jurnal Maranatha*, vol. 1, no. 2, pp. 615–629, 2019.

[3] P. Venkatasubbu and M. Ganesh, "Used Cars Price Prediction using Supervised Learning Techniques," *International Journal of Engineering and Advanced Technology*, vol. 9, no. 1S3, pp. 216–223, 2019.

[4] Ping-Feng Pai and Wen-Chang Wang, "Using Machine Learning Models and Actual Transaction Data for Predicting Real Estate Prices," *Applied Sciences*, vol. 10, no. 17, p. 5832, 2020.

[5] J. A. Abdella, N. Zaki, K. Shuaib, and F. Khan, "Airline ticket price and demand prediction: A survey," *Journal of King Saud University - Computer and Information Sciences*, vol. 33, no. 4, 2019.

[6] E. Jumin, F. B. Basaruddin, Y. Bte. M. Yusoff, S. D. Latif, and A. N. Ahmed, "Solar radiation prediction using boosted decision tree regression model: A case study in Malaysia," *Environmental Science and Pollution Research*, vol. 28, no. 21, pp. 26571–26583, 2021.

[7] H. Lyu, "Cryptocurrency Price forecasting: A Comparative Study of Machine Learning Model in Short-Term Trading," *IEEE Xplore*, 2022.

[8] K. M. Hindrayani, T. M. Fahrudin, R. P. Aji, and E. M. Safitri, "Indonesian Stock Price Prediction including Covid19 Era Using Decision Tree Regression," *IEEE Xplore*, 2020.

[9] CH. R. Madhuri, G. Anuradha, and M. V. Pujitha, "House Price Prediction Using Regression Techniques: A Comparative Study," *IEEE Xplore*, 2019.

[10] M. Thamarai and S. P. Malarvizhi, "House Price Prediction Modeling Using Machine Learning," *International Journal of Information Engineering and Electronic Business*, vol. 12, no. 2, pp. 15–20, 2020.

[11] S. Lahmiri, S. Bekiros, and C. Avdoulas, "A comparative assessment of machine learning methods for predicting housing prices using Bayesian optimization," *Decision Analytics Journal*, vol. 6, p. 100166, 2023.

[12] M. Javed Awan, M. Shafry Mohd Rahim, H. Nobanee, A. Munawar, A. Yasin, and A. Mohd Zain Azlanmz, "Social Media and Stock Market Prediction: A Big Data Approach," *Computers, Materials & Continua*, vol. 67, no. 2, pp. 2569–2583, 2021.

[13] E. Eka Patriya, "IMPLEMENTASI SUPPORT VECTOR MACHINE PADA PREDIKSI HARGA SAHAM GABUNGAN (IHSG)," *Jurnal Ilmiah Teknologi dan Rekayasa*, vol. 25, no. 1, pp. 24–38, 2020.

[14] R. Nopianti, A. T. Panudju, and A. Permana, "Prediksi Harga Saham Indonesia pada Masa Covid-19 Menggunakan Regresi Pohon Keputusan," *Jurnal Ecodemica Jurnal Ekonomi Manajemen dan Bisnis*, vol. 6, no. 1, pp. 68–76, 2022.

[15] O. Somantri, S. Purwaningrum, and R, Riyanto, "MODEL SUPPORT VEKTOR MACHINE (SVM) BERDASARKAN PARAMETER WINDOWS UNTUK PREDIKSI KEKUATAN GEMPA BUMI," *JTT (Jurnal Teknologi Terapan)*, vol. 8 no. 1, pp. 17–24, 2022.

[16] M. A. Hasanah, S. Soim, and A. S. Handayani, "Implementasi CRISP-DM Model Menggunakan Metode Decision Tree dengan Algoritma CART untuk Prediksi Curah Hujan Berpotensi Banjir," *Journal of Applied Informatics and Computing*, vol. 5, no. 2, pp. 103–108, 2021.

[17] A. F. Boy, “Implementasi Data Mining Dalam Memprediksi Harga Crude Palm Oil (CPO) Pasar Domestik Menggunakan Algoritma Regresi Linier Berganda (Studi Kasus Dinas Perkebunan Provinsi Sumatera Utara),” *Journal of Science and Social Research*, vol. 3, no. 2, pp. 78–85, 2020.

[18] Fatkhuroji, S. Santosa, and R. A. Pramunendar, “PREDIKSI HARGA KEDELAI LOKAL DAN KEDELAI IMPOR DENGAN METODE SUPPORT VECTOR MACHINE BERBASIS FORWARD SELECTION,” *Jurnal Cyberku*, vol. 15, no. 1, pp. 61–76, 2019.

[19] D. Sepri and A. Fauzi, “PREDIKSI HARGA CABAI MERAH MENGGUNAKAN SUPPORT VECTOR REGRESSION,” *Computer Based Information System Journal*, vol. 8, no. 2, pp. 1–5, 2020.

[20] A. Saiful, “Prediksi Harga Rumah Menggunakan Web Scrapping dan Machine Learning Dengan Algoritma Linear Regression,” *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, vol. 8, no. 1, pp. 41–50, 2021.

[21] H. W. Herwanto, T. Widiyaningtyas, and P. Indriana, “Penerapan Algoritme Linear Regression untuk Prediksi Hasil Panen Tanaman Padi,” *Jurnal Nasional Teknik Elektro dan Teknologi Informasi (JNTETI)*, vol. 8, no. 4, p. 364, 2019.

[22] A. R. Aziz, B. Warsito, and A. Prahutama, “PENGARUH TRANSFORMASI DATA PADA METODE LEARNING VECTOR QUANTIZATION TERHADAP AKURASI KLASIFIKASI DIAGNOSIS PENYAKIT JANTUNG,” *Jurnal Gaussian*, vol. 10, no. 1, pp. 21–30, 2021.