



## **PERBANDINGAN ALGORITMA RANDOM FOREST REGRESSOR DAN SUPPORT VECTOR MACHINE DALAM PREDIKSI HARGA RUMAH DI KALIMANTAN BARAT**

**Gebby Gisela, Sucipto**

Program Studi Teknik Informatika, Fakultas Teknik dan Ilmu Komputer,

Universitas Muhammadiyah Pontianak

### **Abstrak**

Pertumbuhan industri otomotif di Indonesia turut mendorong tingginya permintaan terhadap mobil bekas sebagai alternatif yang lebih ekonomis dibandingkan mobil baru. Namun, penentuan harga mobil bekas sering kali menjadi tantangan bagi showroom maupun calon pembeli karena melibatkan banyak faktor dan bersifat subjektif. Penelitian ini bertujuan untuk membangun model prediksi harga mobil bekas menggunakan algoritma Support Vector Regression (SVR) dengan pendekatan Radial Basis Function (RBF) kernel. Data sebanyak 1.000 entri diperoleh melalui teknik web scraping dari situs [cintamobil.com](http://cintamobil.com). Metodologi penelitian mengacu pada kerangka kerja CRISP-DM, dimulai dari pemahaman bisnis hingga deployment model melalui aplikasi web menggunakan Streamlit. Proses preprocessing melibatkan penanganan missing value, outlier, duplikasi data, serta transformasi fitur numerik dan kategorikal. Model SVR dievaluasi menggunakan metrik RMSE, MAPE, dan MAE untuk menilai akurasi prediksi. Hasil penelitian menunjukkan bahwa SVR mampu memberikan prediksi harga yang cukup akurat, dengan parameter  $C=1$ ,  $\gamma=0.1$ , dan  $\epsilon=0.1$  yang menghasilkan performa terbaik, yaitu nilai MAE sebesar Rp 6.472.572, RMSE sebesar Rp 8.958.555, dan MAPE sebesar 3,41%. Mengacu pada kategori tingkat akurasi prediksi berdasarkan nilai MAPE, di mana nilai  $MAPE \leq 10\%$  dikategorikan sebagai akurasi tinggi, maka model ini dapat disimpulkan memiliki akurasi prediksi yang tinggi. Hal ini menunjukkan bahwa model SVR yang digunakan mampu memperkirakan harga mobil bekas dengan tingkat kesalahan yang rendah dan ketepatan yang baik.

**Kata Kunci:** Prediksi Harga Rumah, Random Forest Regressor, Support Vector Regression, Machine Learning.

## PENDAHULUAN

Pasar properti di Indonesia menghadapi masalah ketidakseimbangan informasi antara penjual dan pembeli, di mana tidak adanya standar penilaian yang transparan membuat pembeli kesulitan menilai kewajaran harga yang ditawarkan. Model penilaian konvensional yang mengandalkan penilaian subjektif agen properti atau metode perbandingan sederhana seringkali gagal menangkap kompleksitas hubungan antar variabel penentu harga seperti karakteristik fisik bangunan, lokasi strategis, dan dinamika pasar sehingga menghasilkan perkiraan yang tidak akurat dan tidak konsisten.

Sejumlah penelitian terdahulu telah mengkaji prediksi harga rumah menggunakan berbagai pendekatan. Soegianto dkk. (2024) membandingkan beberapa algoritma termasuk Random Forest Regressor dan Support Vector Regression pada data properti bisnis, di mana Random Forest Regressor mencapai nilai  $R^2$  sebesar 0,81 dan menunjukkan kinerja kompetitif dalam menangani data non-linear. Tanamal dkk. (2023) melaporkan akurasi tertinggi sebesar 88% menggunakan Random Forest pada data rumah di Surabaya. Dari sisi SVR, Almahdy & Pradnya (2024) menunjukkan bahwa SVR dengan kernel RBF mampu menghasilkan MAPE sebesar 18,82% dan  $R^2$  sebesar 0,61 pada data properti di Kabupaten Bantul, membuktikan kemampuannya memodelkan hubungan non-linear. Sebagai perbandingan metodologis, Penalun dkk. (2023) secara langsung membandingkan Random Forest Regression dan SVR, menegaskan bahwa keduanya layak diadu dalam tugas regresi serupa.

Meskipun penelitian prediksi harga rumah berbasis machine learning telah berkembang pesat, hampir seluruhnya berfokus pada dataset dari luar negeri atau kota-kota besar di Pulau

Jawa seperti Jabodetabek, Bandung, dan Surabaya. Penerapan algoritma Random Forest Regressor dan SVR pada wilayah Kalimantan Barat masih sangat terbatas, padahal wilayah ini memiliki karakteristik pasar properti yang berbeda dipengaruhi oleh kondisi geografis, pola pembangunan wilayah, serta infrastruktur lokal yang berpotensi memberikan pengaruh berbeda terhadap harga rumah. Selain itu, belum tersedia sistem prediksi harga rumah berbasis aplikasi interaktif yang dapat digunakan langsung oleh masyarakat di wilayah tersebut. Penelitian ini hadir untuk mengisi celah tersebut dengan memanfaatkan dataset hasil web scraping dari Rumah123.com yang mencakup 1.019 entri properti di Kalimantan Barat.

Penelitian ini dilakukan dengan tiga tujuan utama: pertama, membandingkan kinerja algoritma Random Forest Regressor dan Support Vector Regression dalam memprediksi harga rumah di Kalimantan Barat menggunakan data hasil web scraping dari Rumah123.com sebanyak 1.019 entri; kedua, mengevaluasi performa kedua model menggunakan metrik MAE, RMSE,  $R^2$ , dan MAPE; dan ketiga, mengembangkan aplikasi prediksi harga rumah berbasis web menggunakan Streamlit sebagai antarmuka interaktif bagi masyarakat umum, calon pembeli, penjual, maupun pengembang properti. Hasil penelitian menunjukkan bahwa Random Forest Regressor unggul atas SVR dengan nilai MAE Rp699.529.405, RMSE Rp1.477.106.660,  $R^2$  sebesar 0,5779, dan MAPE 39,71%, dibandingkan SVR yang hanya mencapai  $R^2$  0,2563 dan MAPE 46,53%.

## METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen berbasis machine learning untuk memprediksi harga rumah di wilayah Kalimantan Barat. Subjek

penelitian berupa data properti yang diperoleh dari situs Rumah123.com melalui teknik web scraping, sehingga menghasilkan dataset “Harga Rumah Kalimantan Barat” yang merepresentasikan kondisi pasar properti pada periode tertentu.

Alat yang digunakan dalam penelitian ini adalah bahasa pemrograman Python dengan beberapa library pendukung seperti Scikit-learn untuk pemodelan machine learning, Pandas dan NumPy untuk pengolahan data, serta Streamlit untuk implementasi sistem berbasis web. Desain penelitian mengacu pada alur CRISP-DM yang meliputi pemahaman bisnis, pemahaman data, persiapan data, pemodelan, evaluasi, dan deployment sistem.

Teknik pengambilan data dilakukan menggunakan web scraping tanpa sampling manual, sehingga seluruh data yang berhasil dikumpulkan digunakan sebagai dataset penelitian. Variabel yang digunakan terdiri dari variabel independen berupa luas tanah, luas bangunan, jumlah kamar tidur, dan jumlah kamar mandi, sedangkan variabel dependen adalah harga rumah.

Tahapan pengolahan data meliputi preprocessing yang mencakup penanganan missing value, outlier, duplikasi data, serta transformasi fitur numerik dan kategorikal. Model yang digunakan dalam penelitian ini adalah Random Forest Regressor dan Support Vector Regression (SVR) dengan kernel Radial Basis Function (RBF).

Analisis dan evaluasi model dilakukan dengan membandingkan performa kedua algoritma berdasarkan metrik regresi yaitu Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), R-Squared ( $R^2$ ), dan Mean Absolute Percentage Error (MAPE). Hasil evaluasi digunakan untuk menentukan model terbaik dalam memprediksi harga rumah berdasarkan tingkat kesalahan dan akurasi prediksi.

## HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil evaluasi dari Perbandingan Random Forest Regressor dan Support Vector Regressor dalam sistem prediksi harga Rumah. Perbandingan ini dilakukan untuk menilai akurasi model berdasarkan metrik MAE, RMSE,  $R^2$ , dan MAPE. Pembahasan difokuskan pada pemilihan konfigurasi parameter terbaik yang mampu menghasilkan performa prediksi paling optimal.

### Implementasi Model Prediksi

Pada tahap implementasi model prediksi, dilakukan dua proses utama yaitu pemisahan data (train-test split) dan pembangunan model prediksi menggunakan algoritma Random Forest Regressor dan Support Vector Regression. Langkah ini dilakukan untuk memastikan bahwa model dapat dilatih dan diuji secara terstruktur, serta mampu menghasilkan performa prediksi yang optimal berdasarkan konfigurasi pembagian data yang digunakan.

Proses pertama adalah membagi dataset menjadi data latih dan data uji menggunakan fungsi `train_test_split` dari pustaka `scikit-learn`. Pemisahan ini bertujuan menilai kemampuan generalisasi model terhadap data yang belum pernah dilihat sebelumnya. Evaluasi dilakukan terhadap tiga rasio pembagian data, yaitu 60:40, 70:30, dan 80:20. Berikut merupakan kode pemisahan data yang digunakan. Berdasarkan kode di atas, dilakukan evaluasi model dengan menggunakan beberapa rasio pembagian data, sebagaimana ditampilkan pada Tabel 3.1 berikut.

**Tabel 1. Tabel Barang**

| Rasio | RMSE          | MAPE  | MAE         |
|-------|---------------|-------|-------------|
| 60:40 | 1.477.106.660 | 44,12 | 699.529.405 |
| 70:30 | 1.960.718.420 | 46,53 | 861.454.906 |

|       |               |       |             |
|-------|---------------|-------|-------------|
| 80:20 | 1.477.106.660 | 39,71 | 699.529.405 |
|-------|---------------|-------|-------------|

Dari hasil evaluasi pada Tabel 3.1, rasio 80:20 memberikan performa terbaik dengan nilai RMSE, MAPE, dan MAE yang paling rendah dibandingkan rasio lainnya. Rasio ini menunjukkan bahwa penggunaan proporsi data latih yang lebih besar memungkinkan model mempelajari pola data lebih optimal sehingga menghasilkan generalisasi yang lebih stabil saat diuji pada data baru. Oleh karena itu, rasio 80% data latih dan 20% data uji dipilih sebagai konfigurasi optimal dalam penelitian ini.

Tahap berikutnya adalah membangun model regresi menggunakan Random Forest Regressor dan Support Vector Regression. Sebelum proses pelatihan, dilakukan transformasi logaritmik pada variabel target (harga rumah) untuk mengurangi pengaruh outlier dan menstabilkan distribusi data yang memiliki rentang nilai sangat besar. Pada model Random Forest Regressor, model diinisialisasi menggunakan kelas RandomForestRegressor dari pustaka scikit-learn dengan menentukan jumlah pohon ( $n_{\text{estimators}}$ ) dan nilai  $\text{random\_state}$  untuk menjaga konsistensi hasil eksperimen. Sementara itu, model SVR dilatih menggunakan kelas SVR dari pustaka scikit-learn disertai proses hyperparameter tuning untuk memperoleh kombinasi parameter yang lebih optimal. Berikut adalah kode pelatihan kedua model yang digunakan.

Setelah proses pelatihan selesai, kedua model digunakan untuk melakukan prediksi terhadap data uji. Hasil prediksi yang masih berada dalam skala logaritmik kemudian dikembalikan ke skala aslinya menggunakan fungsi `inverse transform (np.expm1)` untuk keperluan evaluasi.

### Hasil Evaluasi Random Forest Regressor

Setelah mendapatkan rasio pembagian data terbaik, tahap selanjutnya adalah mengevaluasi kinerja model Random Forest Regressor menggunakan data uji. Evaluasi dilakukan menggunakan empat metrik, yaitu Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), R-Squared ( $R^2$ ), dan Mean Absolute Percentage Error (MAPE). Hasil evaluasi disajikan pada Tabel 3.2 berikut.

**Tabel 3.2. Hasil Evaluasi Random Forest Regressor**

| Metrik Evaluasi                       | Nilai            |
|---------------------------------------|------------------|
| Mean Absolute Error (MAE)             | Rp 699.529.405   |
| Root Mean Squared Error (RMSE)        | Rp 1.477.106.660 |
| R-Squared ( $R^2$ )                   | 0,5779           |
| Mean Absolute Percentage Error (MAPE) | 39,71%           |

Berdasarkan Tabel 3.2, nilai MAE sebesar Rp699.529.405 menunjukkan bahwa rata-rata kesalahan prediksi model berada pada kisaran ratusan juta rupiah, yang mencerminkan variasi harga properti yang sangat luas pada data penelitian ini. Nilai RMSE yang lebih besar mengindikasikan adanya beberapa kesalahan prediksi yang cukup tinggi, terutama pada rumah dengan harga ekstrem. Nilai  $R^2$  sebesar 0,5779 menunjukkan bahwa model mampu menjelaskan sekitar 57,79% variasi harga rumah, sedangkan nilai MAPE sebesar 39,71% masih tergolong dalam kategori wajar berdasarkan standar interpretasi MAPE (MAPE < 10% = sangat baik, 10–20% = baik, 20–50% = wajar, >50% = buruk) mengingat rentang harga properti yang sangat bervariasi.

### Hasil Evaluasi Model Support Vector Regression

Evaluasi model Support Vector Regression (SVR) dilakukan menggunakan metrik yang sama dengan

model Random Forest Regressor untuk memastikan perbandingan yang objektif. Hasil evaluasi model SVR disajikan pada Tabel 3.3 berikut.

**Tabel 3.2. Hasil Evaluasi Support Vector Regression**

| Metrik Evaluasi                       | Nilai            |
|---------------------------------------|------------------|
| Mean Absolute Error (MAE)             | Rp 861.454.906   |
| Root Mean Squared Error (RMSE)        | Rp 1.960.718.420 |
| R-Squared ( $R^2$ )                   | 0,2563           |
| Mean Absolute Percentage Error (MAPE) | 46,53%           |

Berdasarkan Tabel 3.3, model SVR memiliki nilai MAE sebesar Rp861.454.906 dan RMSE sebesar Rp1.960.718.420 yang keduanya lebih besar dibandingkan Random Forest Regressor. Nilai  $R^2$  sebesar 0,2563 menunjukkan bahwa model SVR hanya mampu menjelaskan sekitar 25,63% variasi harga rumah, yang mengindikasikan kemampuan generalisasi model yang lebih terbatas. Nilai MAPE sebesar 46,53% juga masih dalam kategori wajar namun mendekati batas atas kategori tersebut. Meskipun model SVR telah mengalami peningkatan performa setelah dilakukan transformasi logaritmik dan penyesuaian hyperparameter, model ini masih memiliki keterbatasan dalam menangani data dengan rentang harga yang sangat luas dan distribusi yang tidak merata.

### Perbandingan Kinerja Model

Untuk mempermudah analisis perbandingan kinerja kedua model, hasil evaluasi Random Forest Regressor dan Support Vector Regression dirangkum dalam Tabel 3.4 berikut.

**Tabel 3.4. Tabel Perbandingan Kinerja Model Random Forest Regressor dan SVR**

| Model                     | MAE(Rp)     | RMSE(Rp)      | $R^2$ (%) | MAPE(%) |
|---------------------------|-------------|---------------|-----------|---------|
| Random Forest Regressor   | 699.529.405 | 1.477.106.660 | 0,5779    | 39,71   |
| Support Vector Regression | 861.454.906 | 1.960.718.420 | 0,2563    | 46,53   |

Berdasarkan Tabel 3.4, dapat dilihat bahwa Random Forest Regressor menghasilkan nilai MAE dan RMSE yang lebih rendah serta nilai  $R^2$  yang lebih tinggi dibandingkan Support Vector Regression pada seluruh metrik evaluasi. Hal ini menunjukkan bahwa Random Forest Regressor memiliki tingkat akurasi yang lebih baik dan mampu menjelaskan variasi harga rumah dengan lebih optimal.

Keunggulan Random Forest Regressor disebabkan oleh karakteristiknya sebagai metode ensemble learning berbasis pohon keputusan yang efektif dalam menangani hubungan non-linear dan data dengan variasi harga yang besar, yang merupakan karakteristik utama data properti. Dengan membangun sejumlah pohon keputusan secara paralel dan menggabungkan hasilnya, model ini mampu menghasilkan prediksi yang lebih stabil dan robust terhadap nilai ekstrem. Sementara itu, Support Vector Regression meskipun telah mengalami peningkatan performa setelah dilakukan transformasi logaritmik dan penyesuaian parameter, masih memiliki keterbatasan dalam menangani data properti dengan rentang harga yang sangat luas.

Besarnya nilai error yang dihasilkan oleh kedua model dipengaruhi oleh beberapa faktor, antara lain: karakteristik data harga rumah yang memiliki rentang nilai sangat luas sehingga selisih prediksi yang relatif kecil dapat menghasilkan error absolut yang besar dalam satuan rupiah; sebaran data yang tidak merata dan keberadaan nilai ekstrem pada rumah dengan harga tinggi yang meningkatkan sensitivitas RMSE; serta keterbatasan fitur yang digunakan di mana beberapa faktor penting penentu harga rumah belum sepenuhnya terakomodasi dalam dataset. Oleh karena itu, nilai error yang dihasilkan pada penelitian ini masih

dapat diterima dan mencerminkan kondisi realistis data properti di wilayah Kalimantan Barat. Berdasarkan hasil perbandingan tersebut, Random Forest Regressor ditetapkan sebagai model terbaik dalam penelitian ini.

## SIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa algoritma Random Forest Regressor dan Support Vector Regression berhasil diPerbandingkan untuk memprediksi harga rumah di wilayah Kalimantan Barat menggunakan data properti hasil scraping dari platform Rumah123. Proses pemodelan dilakukan dengan memanfaatkan fitur jumlah kamar tidur, jumlah kamar mandi, jumlah garasi, luas bangunan, luas tanah, dan lokasi rumah sebagai variabel prediktor harga rumah. Hasil evaluasi kinerja model menunjukkan bahwa Random Forest Regressor memiliki performa yang lebih baik dibandingkan Support Vector Regression. Hal ini dibuktikan dengan nilai MAE sebesar Rp699.529.405, RMSE sebesar Rp1.477.106.660,  $R^2$  sebesar 0,5779, dan MAPE sebesar 39,71% pada Random Forest Regressor, yang seluruhnya lebih unggul dibandingkan Support Vector Regression dengan MAE sebesar Rp861.454.906, RMSE sebesar Rp1.960.718.420,  $R^2$  sebesar 0,2563, dan MAPE sebesar 46,53%. Random Forest Regressor mampu menjelaskan sekitar 57,79% variasi harga rumah, sementara Support Vector Regression hanya mampu menjelaskan sekitar 25,63%. Dengan demikian, Random Forest Regressor dinilai lebih akurat dan lebih sesuai untuk karakteristik data harga rumah yang memiliki variasi tinggi dan hubungan non linear antarfitur. Selain itu, penelitian ini juga berhasil merancang dan mengembangkan aplikasi prediksi harga rumah berbasis web menggunakan Streamlit. Aplikasi tersebut mampu mengintegrasikan

model prediksi dan menyajikan hasil estimasi harga rumah secara interaktif berdasarkan input karakteristik properti yang diberikan oleh pengguna. Keberadaan aplikasi ini diharapkan dapat membantu pengguna dalam memperoleh gambaran harga rumah secara cepat dan objektif di wilayah Kalimantan Barat.

## DAFTAR PUSTAKA

- Suharyanto, A. (2016). Pusat Aktivitas Ritual Kepercayaan Parmalim di Huta Tinggi Laguboti, *Jurnal Ilmu Pemerintahan dan Sosial Politik UMA*, 4 (2): 182-195.
- L. Li and K. W. Chau, "Information Asymmetry with Heterogeneous Buyers and Sellers in the Housing Market," *Journal of Real Estate Finance and Economics*, vol. 68, no. 1, pp. 138-159, Jan. 2024, doi: 10.1007/s11146-023-09939-y.
- S. An, Y. Song, H. Jang, and K. Ahn, "Toward transparent and accurate housing price appraisal: Hedonic price models versus machine learning algorithms," *Financial Innovation*, vol. 11, no. 1, Dec. 2025, doi: 10.1186/s40854-025-00874-w.
- L. M. Soegianto, A. T. Hinandra, P. A. Suri, and M. Fajar, "Comparison of Model Performance on Housing Business Using Linear Regression, Random Forest Regressor, SVR, and Neural Network," in *Procedia Computer Science*, Elsevier B.V., 2024, pp. 1139-1145. doi: 10.1016/j.procs.2024.10.343.
- R. Tanamal, N. Minoque, T. Wiradinata, Y. Soekamto, and T. Ratih, "House Price Prediction Model Using Random Forest in Surabaya City," *TEM Journal*, vol. 12, no. 1, pp. 126-132, Feb. 2023, doi: 10.18421/TEM121-17.
- R. Fauzan Almahdy and W. D. Mega Pradnya, "Prediksi Harga Rumah Di Kabupaten Bantul Menggunakan Algoritma Support Vector Regression," vol. 11, no. 2, pp. 152-165, 2024, [Online]. Available: <http://jurnal.mdp.ac.id/jatiasi@mdp.ac.idy10th.2012>
- F. E. Penalun, A. Hermawan, and D. Avianto, "Perbandingan Random Forest Regression dan Support Vector Regression Pada Prediksi Laju Penguapan".

“Implementasi Algoritma Random Forest Regression dalam Sistem Prediksi Harga Rumah di Jabodetabek”.

I. P. Putra, I. Ketut, G. Suhartana, and B. Jimbaran, “Perbandingan Akurasi Algoritma Regresi Linier, Regresi Polinomial, dan Support Vector Regression Pada Model Sistem Prediksi Harga Rumah,” 2022. [Online]. Available: <https://www.kaggle.com/datasets/wisnuanggara/daftar-hargarumah>

I. Nabillah and I. Ranggadara, “Mean Absolute Percentage Error untuk Evaluasi Hasil Prediksi Komoditas Laut,” JOINS (Journal of Information System), vol. 5, no. 2, pp. 250–255, Nov. 2020, doi: 10.33633/joins.v5i2.3900.

T. O. Hodson, “Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not,” Jul. 19, 2022, Copernicus GmbH. doi: 10.5194/gmd-15-5481-2022.

N. Mohamad Iqbal Jauhari, R. Wulanningrum, and A. Bagus Setiawan, “Sistem Deteksi Kendaraan Menggunakan StreamLit Metode Yolo Universitas Nusantara PGRI Kediri,” Online, 2024.