JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 10, No. 4, December 2025, Pp. 3301-3312

ESTIMASI HARGA PROPERTI RUMAH MENGGUNAKAN ALGORITMA MULTIPLE LINEAR REGRESSION DAN RANDOM FOREST REGRESSOR DENGAN TEKNIK WEB SCRAPING PADA WEBSITE PENJUALAN RUMAH DI YOGYAKARTA

Rahmat Sobirin*[1)], Kusrini[2)], Ferry Wahyu Wibowo[3)]
1. Magister Informatika, Universitas Amikom Yogyakarta, Indonesia
2. Magister Informatika, Universitas Amikom Yogyakarta, Indonesia
3. Magister Informatika, Universitas Amikom Yogyakarta, Indonesia

Article Info
Kata Kunci: *Multiple Linear Regression; Random Forest Regressor; Web Scraping;* Estimasi Harga; Properti Rumah.

Keywords: *Linear Regression; Random Forest Regressor; Web Scraping; Price Estimation; House Property.*

Article history:
Received 28 August 2024
Revised 22 October 2024
Accepted 28 October 2024
Available online 1 December 2025

DOI :
https://doi.org/10.29100/jipi.v10i4.6463

* Corresponding author.
Rahmat Sobirin
E-mail address:
rahmat.sob@students.amikom.ac.id

ABSTRAK
Penelitian ini mengeksplorasi estimasi harga properti rumah di Yogyakarta dengan menggunakan algoritma Multiple Linear Regression (MLR) dan Random Forest Regressor (RFR), didukung oleh teknik web scraping untuk pengumpulan data. Teknik web scraping digunakan untuk mengumpulkan data harga rumah dari situs penjualan online, mencakup informasi seperti lokasi, luas bangunan, dan fitur-fitur lainnya. Data yang terkumpul diproses dan dianalisis dengan Algoritma *MLR* dan *RFR* untuk menghasilkan model estimasi harga yang akurat. Hasil analisis menunjukkan bahwa kedua algoritma efektif dalam memprediksi harga rumah, dengan Random Forest Regressor memberikan hasil yang sedikit lebih baik dengan nilai R-Square sebesar 0.943 atau 94,3% dibandingkan Multiple Linear Regression dengan nilai R-Square sebesar 0.90 atau 90%. Penelitian ini menyoroti potensi kombinasi teknik web scraping dan machine learning dalam meningkatkan keakuratan estimasi harga properti serta memberikan wawasan berharga bagi pengembangan model prediksi harga di pasar real estat lokal.

ABSTRACT
This research explores house property price estimation in Yogyakarta using Multiple Linear Regression (MLR) and Random Forest Regressor (RFR) algorithms, supported by web scraping techniques for data collection. Web scraping was employed to gather house price data from online sales websites, including information such as location, building size, and other features. The collected data was processed and analyzed using MLR and RFR algorithms to produce accurate price estimation models. The analysis results show that both algorithms are effective in predicting house prices, with the Random Forest Regressor performing slightly better, yielding an R-Square value of 0.943 or 94.3%, compared to Multiple Linear Regression with an R-Square value of 0.90 or 90%. This study highlights the potential of combining web scraping techniques and machine learning in improving the accuracy of property price estimations and provides valuable insights for developing price prediction models in the local real estate market.

.

I. PENDAHULUAN

RUMAH merupakan kebutuhan mendasar bagi masyarakat yang tidak bisa diabaikan, karena berperan sebagai tempat tinggal dan istirahat setelah menjalani rutinitas harian. Selain menjadi kebutuhan primer rumah juga berfungsi sebagai bentuk investasi jangka panjang, mirip dengan emas, karena nilainya cenderung berfluktuasi seiring waktu terutama permintaan terhadap rumah terus meningkat di lokasi yang strategis yang tentunya mempengaruhi harga rumah secara signifikan [1]. Di Indonesia, harga rumah memiliki karakteristik unik yang berbeda dari negara lain. Perbedaan ini dipengaruhi oleh berbagai faktor lokal, seperti pertumbuhan populasi, peningkatan biaya konstruksi, naiknya permintaan akan rumah, serta pertumbuhan PDB. Faktor lain yang memengaruhi harga rumah meliputi kualitas fisik, akses lokasi, karakteristik desain, fasilitas, dan lainnya [2]. Perbedaan karakteristik harga rumah di Indonesia, baik dibandingkan dengan negara lain maupun antar kota dalam

negeri, menciptakan beragam tantangan dan peluang bagi penelitian, serta memberikan kontribusi penting bagi pengetahuan dan praktik di bidang properti. Situasi ini akan mendorong masyarakat yang ingin membeli rumah untuk menimbang apakah properti yang akan dibeli memiliki peluang keuntungan yang baik atau tidak, terutama karena harga rumah terus mengalami kenaikan seiring berjalannya waktu [3]. Sebagai contoh, wilayah Yogyakarta - Indonesia dijadikan studi kasus dalam penelitian ini untuk memprediksi atau mengestimasi harga properti rumah dengan menerapkan teknik web scraping menggunakan Algoritma *Multiple Linear Regression* dan *Random Forest Regressor*. Pemilihan Yogyakarta sebagai lokasi studi dalam penelitian ini didasarkan pada beberapa alasan diantaranya Investasi dalam sektor properti di Yogyakarta semakin meningkat seiring dengan perkembangan, membuatnya menjadi lokasi yang dinamis untuk studi harga properti, dimana tingkat investasi di kabupaten Sleman dan Kota Yogyakarta lebih mendominasi [4]. Jogja juga sebagai salah satu destinasi wisata utama di Indonesia, memiliki daya tarik budaya dan pariwisata yang kuat. Yogyakarta memiliki karakter sosial dan budaya yang kuat dengan adat istiadat yang masih dipegang teguh. Hal ini memengaruhi preferensi dan pola pembelian properti masyarakat lokal, yang bisa berbeda dari kota lain.

Web Scraping telah menjadi teknik yang banyak digunakan untuk secara otomatis dan efisien mengumpulkan data dari internet. Dengan kata lain, daripada menyalin dan menempelkan informasi secara manual dari situs web ke dalam spreadsheet, web scraping memungkinkan aplikasi komputer untuk melakukannya dengan lebih akurat dan jauh lebih cepat dibandingkan dengan manusia [5]. Melalui web scraping, kita dapat mengumpulkan data dan informasi tentang properti rumah dari berbagai sumber online. Data yang dikumpulkan kemudian dapat digunakan sebagai input untuk model pembelajaran mesin dalam memperkirakan harga properti rumah. Estimasi merupakan proses perkiraan atau prediksi yang berkaitan dengan populasi dan sampel melalui teknik pendugaan yang tepat. Estimasi dapat dibagi menjadi empat jenis, yaitu pendugaan titik, pendugaan interval, pendugaan rata-rata, dan pendugaan proporsi [6]. Dalam konteks harga rumah, estimasi digunakan untuk memprediksi nilai properti pada saat ini (waktu eksisting) atau di masa mendatang. Dengan menggunakan model pembelajaran mesin seperti Algoritma *Multiple Linear Regression dan Random Forest Regressor*, data dalam jumlah besar dapat diolah dan berbagai fitur atau variabel yang mempengaruhi harga properti dapat dianalisis. Hal ini membuat estimasi harga yang dihasilkan lebih mendekati nilai pasar sebenarnya.

Penelitian ini memberikan kontribusi unik dalam beberapa aspek yang membedakannya dari penelitian sebelumnya yang menggunakan algoritma serupa, yaitu *Multiple Linear Regression* dan *Random Forest Regressor*, khususnya dalam konteks estimasi harga properti rumah. Penelitian ini memanfaatkan teknik web scraping untuk mengumpulkan data harga properti secara langsung dari website penjualan rumah di Yogyakarta. Penelitian ini secara khusus menyoroti pasar properti di Yogyakarta, yang memiliki karakteristik dan dinamika pasar tersendiri, pada penelitian ini tidak hanya menggunakan variabel umum seperti luas tanah, jumlah kamar, dan lokasi, tetapi juga menambahkan beberapa variabel-variabel unik yang relevan dengan kondisi lokal Yogyakarta. Meskipun *Multiple Linear Regression* dan *Random Forest Regressor* telah digunakan dalam penelitian lain, penelitian ini memberikan analisis komparatif di antara kedua algoritma tersebut dalam konteAks data real-time dari Yogyakarta.

Model machine learning, seperti *Multiple Linear Regression* dan *Random Forest Regression*, telah terbukti efektif dalam berbagai penelitian yang bertujuan untuk memprediksi harga properti. Model-model ini sering kali digunakan karena kemampuannya dalam menangani data yang kompleks dan memberikan hasil yang memuaskan. Misalnya, penelitian yang dilakukan [7] mengeksplorasi penggunaan Multiple Linear Regression dan Random Forest Regression dalam konteks prediksi harga properti. Hasil dari penelitian ini menunjukkan bahwa kedua model tersebut mampu menghasilkan prediksi harga yang cukup akurat. Namun, algoritma Random Forest Regression muncul sebagai yang paling unggul, dengan tingkat akurasi mencapai 81,6%. Temuan ini memperkuat posisi Random Forest Regression sebagai algoritma yang handal dan efektif dalam memprediksi harga properti. Pada Konteks lokasi dan pasar properti yang menjadi objek penelitian tidak dijelaskan secara spesifik dan penelitian ini menggunakan data dari sumber kaggle.com . Adapun pada penelitian ini memiliki keterbatasan fitur atau variabel serta data yang digunakan sedangkan pada penelitian penulis menambahkan beberapoa fitur lainnya.

Selanjutnya, penelitian yang dilakukan [8] memprediksi harga rumah di Pusat Kota Artvin, Turki, menggunakan algoritma Machine Learning. Hasil penelitian ini menunjukkan bahwa algoritma XGBoost dan Random Forest (RF) memberikan performa terbaik dalam memprediksi nilai properti, berdasarkan kriteria seperti Korelasi R^2 (R-Square) dengan masing-masing skor sebesar 0,705 dan 0,701. Penelitian ini memiliki keterbatasan utama adalah jumlah data, meskipun jumlah data yang digunakan terbatas algoritma tersebut berhasil menghasilkan prediksi yang memuaskan. Pada penelitian penulis meskipun teknik web scraping memungkinkan pengumpulan data yang lebih

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 10, No. 4, December 2025, Pp. 3301-3312

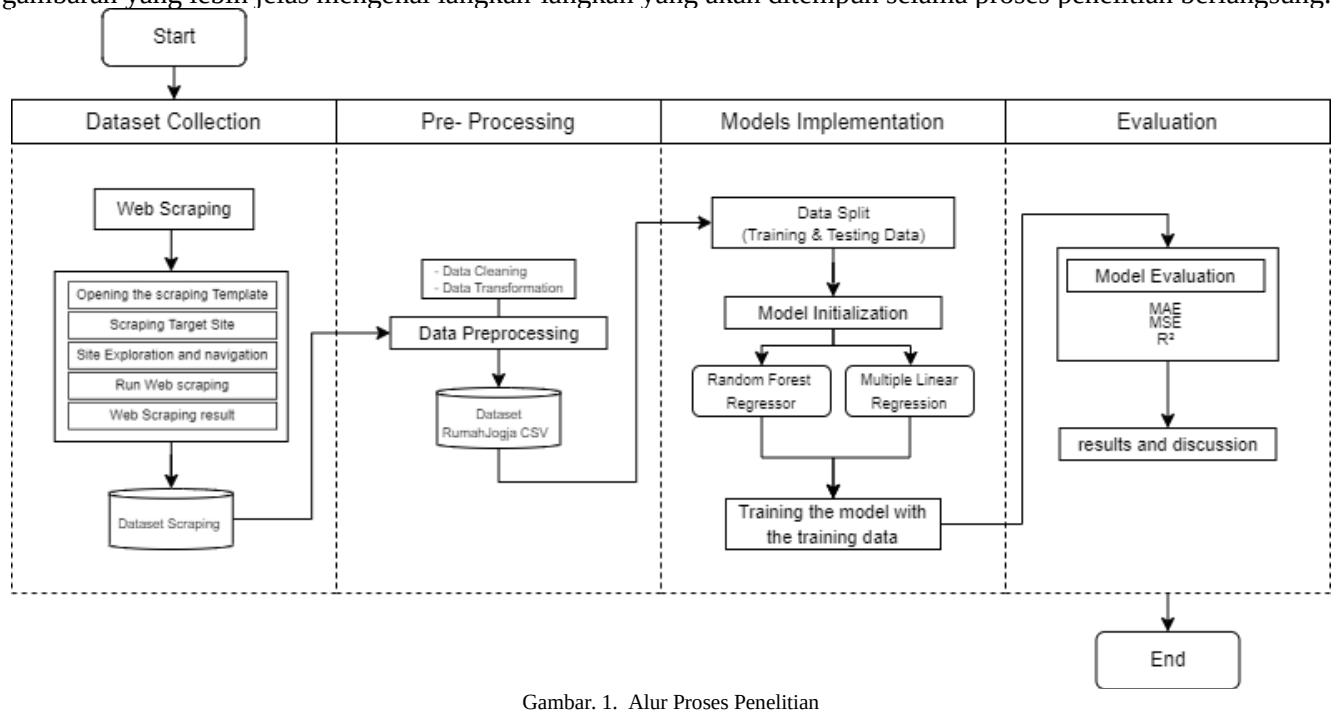
banyak, kualitas dan kelengkapan data dapat menjadi masalah, terutama jika data yang diperoleh dari website memiliki inkonsistensi atau missing values.

Penelitian yang dilakukan oleh [9] memfokuskan pada penggunaan model Machine Learning untuk memprediksi harga rumah di kawasan KLCC, Kuala Lumpur. Studi ini melibatkan analisis harga rumah dengan menggunakan berbagai model seperti multiple regression analysis, Ridge Regression, LightGBM, dan XGBoost. Penelitian ini juga menekankan pentingnya persiapan data yang baik, termasuk proses pembersihan dan transformasi dataset. Selain itu, penelitian ini menyoroti perlunya menambahkan lebih banyak atribut dan memperluas cakupan penelitian ke lokasi lain, serta penggunaan data yang lebih relevan untuk meningkatkan akurasi prediksi. Penelitian ini terbatas pada lokasi di Kuala Lumpur dan menggunakan dataset yang telah ada dengan jumlah variabel terbatas. Adapun perbandingan pada penelitian ini seperti perlu adanya eksplorasi dataset terutama dari segi harga Rumah serta variabel lainnya yang lebih realtime / terbaru. Begitupun pada penlitian penulis mengeksplorasi model machine leraning seperti yaitu *Multiple Linear Regression* dan *Random Forest Regressor*.

Penelitian oleh [10] menerapkan algoritma machine learning untuk memprediksi harga properti di Hong Kong, menggunakan data seperti luas lantai, usia, dan lokasi. Algoritma yang diuji termasuk Support Vector Machine (SVM), Random Forest (RF), dan Gradient Boosting Machine (GBM), serta dilakukan analisis statistik deskriptif dan korelasi. Hasil penelitian menunjukkan bahwa model-model tersebut memberikan prediksi yang akurat dan menunjukkan potensi besar machine learning dalam penilaian nilai properti. Adapun saran dari penelitian ini dapat mengeksplorasi penggunaan data tambahan dan jenis properti lainnya untuk meningkatkan akurasi prediksi harga properti. Perbandingan pada penelitian ini diantaranya Data yang digunakan pada penelitian sebelumnya meliputi atribut seperti luas lantai, usia, dan lokasi properti sedangkan pada penelitian penulis menambahkan atribut tambahan dimana pada penelitian penulis kemungkinan besar menambahkan atribut yang tersedia pada listing website seperti harga, lokasi, ukuran, jumlah kamar, dan fasilitas lainnya dengan metode pengumpulan data menggunakan teknik webs craping.

II. METODE PENELITIAN

Penelitian ini merupakan penelitian kuantitatif yang menggunakan pendekatan eksperimental dan studi kasus. Data yang digunakan dalam penelitian ini berupa data numerik yang diperoleh dari hasil web scraping dari situs web penjualan rumah di wilayah Yogyakarta. Data tersebut kemudian dianalisis dengan menggunakan teknik statistik serta model Algoritma *Multiple Linear Regression* dan *Random Forest Regressor* untuk memperkirakan harga properti rumah. Agar mempermudah pemahaman mengenai penelitian ini, setiap tahapan yang akan dilakukan dijelaskan secara terperinci seperti yang ditampilkan pada Gambar 1. Penjabaran ini bertujuan untuk memberikan gambaran yang lebih jelas mengenai langkah-langkah yang akan ditempuh selama proses penelitian berlangsung.

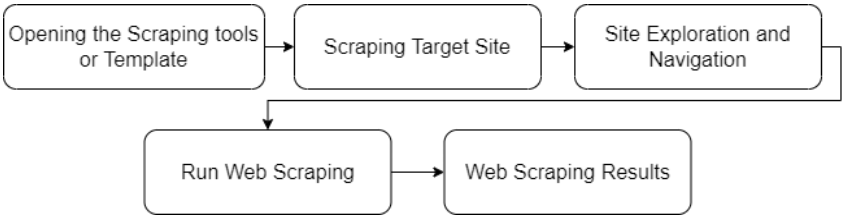


Gambar. 1. Alur Proses Penelitian

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 10, No. 4, December 2025, Pp. 3301-3312

A. *Web Scraping*

Web scraping adalah sebuah teknik yang digunakan untuk mengambil atau mengekstrak data dari situs web dengan memanfaatkan perangkat lunak komputer. Metode ini dirancang untuk mengolah data web yang awalnya tidak terstruktur, seperti informasi yang tersebar di halaman web dalam format yang tidak konsisten atau tidak terorganisir, dan mengubahnya menjadi format yang terstruktur dan sistematis. Data yang telah diolah dapat disimpan dalam basis data atau spreadsheet, sehingga memudahkan untuk dianalisis dan digunakan dalam berbagai aplikasi analitik dan pengolahan data [11]. Adapun proses dalam melakukan *web scraping* dibagi menjadi beberapa tahapan [12], seperti yang ditunjukkan pada Gambar 2.



Gambar. 2. Tahapan-tahapan web scraping

Dimulai tahapan awal yaitu Membuka *Template* atau Tools yang digunakan dalam melakukan proses web scraping seperti Aplikasi Jupyter Notebook atau Google Colaboratory dapat digunakan bersama dengan bahasa pemrograman Python dan library BeautifulSoup. BeautifulSoup, sebuah library Python dimanfaatkan untuk mengekstrak, memeriksa, serta memodifikasi data dari pohon Document Object Model (DOM) pada situs web yang didasarkan pada mesin analisis Hypertext Markup Language (HTML) dan Extensible Markup Language (XML) [13]. Selain itu juga bisa menggunakan beberapa Toolos lain seperti Tools yang disediakan Ekstensi dari Browser Chrome.

Tahapan kedua yaitu Menargetkan Situs atau website yang akan dilakukan scraping. Pada penelitian ini menggunakan sumber website penjualan properti rumah yang ada diyogyakarta seperti dari sumber website *www.rumah123.com* yang telah menyediakan sumber data penjualan rumah beserta fitur dan variabel yang dibutuhkan. Selanjutnya mengeksplorasi situs web yang akan di-scrap untuk memahami fitur atau variabel jual beli properti rumah yang relevan. Gunakan fitur "Inspect" pada browser untuk melihat kode HTML dan CSS halaman tersebut, lalu identifikasi elemen penting dan informasi yang ingin diambil. Setelah itu, masukkan elemen-elemen ini ke dalam template scraping. Setelah semua elemen dimasukkan dan penyesuaian selesai, proses web scraping dimulai dengan menjalankan kode pada template yang telah disiapkan. Hasil scraping akan muncul sesuai dengan elemen-elemen yang telah ditentukan dari situs web tersebut. Kemudian, hasil web scraping ini akan diekspor ke dalam file CSV.

B. Fitur dan variabel Rumah

Fitur dan variabel ini membantu dalam menentukan nilai pasar rumah dengan mempertimbangkan berbagai aspek yang mempengaruhi harga. Fitur atau variabel yang digunakan dalam penelitian ini diantaranya Harga Rumah, Luas Lahan, Luas Bnagunan, Kamar Tidur, Kamar mandi, Carport, Lantai, Tahun, dan Daya Listrik. Untuk Lebih detail dapat dilihat pada tabel 1 Fitur dan Variabel Rumah.

TABEL I
FITUR DAN VARIABEL RUMAH

No	FITUR/VARIABEL	JENIS VARIABEL	Keterangan
1	Harga Rumah	Dependen (Y)	Nilai moneter yang mencerminkan biaya atau harga yang diminta atau ditawarkan untuk membeli sebuah rumah merupakan aspek penting dalam transaksi properti. Harga rumah biasanya ditentukan oleh sejumlah faktor yang saling berinteraksi.
2	Luas Lahan	Independent (X1)	Merujuk pada dimensi area lahan atau tanah tempat rumah dibangun. Ukuran tanah umumnya dinyatakan dalam satuan luas seperti meter persegi (m^2).

3	Luas Bangunan	Independent (X2)	Ukuran total area dari bangunan atau rumah, mencakup seluruh ruang di dalamnya. Luas bangunan biasanya diukur dalam satuan seperti meter persegi (m²).
4	Kamar Tidur	Independent (X3)	Total jumlah kamar yang ada di rumah yang berfungsi sebagai ruang tidur. Jumlah kamar tidur sering digunakan untuk menilai kapasitas hunian dan memenuhi kebutuhan keluarga.
5	Kamar Mandi	Independent (X4)	Jumlah ruangan yang memiliki fasilitas mandi, seperti wastafel, toilet, dan shower. Banyaknya kamar mandi sering kali menjadi pertimbangan utama dalam menilai tingkat kenyamanan dan fungsionalitas rumah.
6	Carport	Independent (X5)	Carport atau parkir adalah area yang dirancang khusus untuk menyimpan kendaraan seperti mobil, mengacu pada jumlah di mana kendaraan dapat diparkir, baik itu garasi, carport, atau area parkir terbuka.
7	Lantai	Independent (X6)	Jumlah lantai merujuk pada total tingkat atau level yang terdapat dalam sebuah bangunan rumah. Sebagai contoh, rumah yang memiliki dua lantai berarti terdapat dua tingkat yang dapat digunakan untuk ruang tinggal atau aktivitas lainnya.
8	Tahun Bangun/Renovasi	Independent (X7)	Merujuk pada tahun di mana sebuah rumah pertama kali dibangun atau tahun terakhir kali rumah tersebut mengalami renovasi.
9	Daya Listrik	Independent (X8)	Kapasitas maksimum sistem listrik di sebuah rumah untuk menyediakan energi listrik, biasanya diukur dalam satuan watt (W)

C. *Preprocessing Data*

Data yang telah dikumpulkan welalui webscraping pada website *www.rumah123.com* kemudian akan melalui tahap pra-pemrosesan, yang merupakan langkah penting dalam memastikan kualitas dan integritas data sebelum analisis lebih lanjut. Proses ini melibatkan identifikasi dan penghapusan titik data yang tidak lengkap, termasuk data yang mengandung nilai NaN atau data yang tidak stabil dan tidak dapat diandalkan [14]. Selain itu, data yang dianggap tidak relevan atau tidak berkaitan langsung dengan tujuan penelitian akan dihilangkan, sehingga hanya data yang relevan dan sesuai dengan konteks penelitian yang akan disimpan dan digunakan [15]. Tahap ini bertujuan untuk membersihkan dataset agar analisis yang dilakukan dapat menghasilkan hasil yang akurat dan valid. Adapaun hasil data sebelum dan setelah dipreprocesing dapat dilihat pada Gambar 3 & 4.

	nama_rumah	harga	lahan	bangunan	KT	KM	carport	Lantai	tahun	listrik	lokasi	link-href
0	Hunian Kolam Renang 300 m ke Jl Jogja Solo Cla...	Rp 2,52 Miliar	171 m²	130 m²	4	3	2.0	2.0	2024.0	3500.0	Kalasan, Sleman	https://www.rumah123.com/properti/sleman/hos17...
1	Rumah murah dekat bandara adisucipto dekat jl ...	Rp 450 Juta	85 m²	48 m²	2	1	1.0	1.0	2023.0	2200.0	Berbah, Sleman	https://www.rumah123.com/properti/sleman/hos17...
2	RUMAH SIAP HUNI SHM DALAM CLUSTER DEKAT KADISO...	Rp 775 Juta	106 m²	80 m²	2	2	1.0	1.0	2024.0	1300.0	Sleman, Sleman	https://www.rumah123.com/properti/sleman/hos17...
3	Dijual Rumah 2 Lantai dan 2,5 Lantai, Baru, De...	Rp 2,43 Miliar	70 m²	130 m²	4	2	1.0	2.0	2024.0	3500.0	Umbulharjo, Yogyakarta	https://www.rumah123.com/properti/yogyakarta/h...
4	Rumah Baru Legalitas Terjamin di Seyegan Slema...	Rp 330 Juta	60 m²	60 m²	2	1	1.0	1.0	2024.0	1300.0	Seyegan, Sleman	https://www.rumah123.com/properti/sleman/hos17...

Gambar. 3. Data Sebelum Preprocesing

	nama_rumah	harga	lahan	bangunan	kamar_tidur	kamar_mandi	carport	Lantai	tahun	listrik	kode_pos	lokasi	link-href
0	RUMAH MEWAH MURAH DALAM CLUSTER PREMIUM DENGAN...	1850000	118	126	4	3	1	2	2024	4400	55282	Maguwoharjo, Sleman	https://www.rumah123.com/properti/yogyakarta/h...
1	Rumah Mewah Di Jual Di Sleman Yogyakarta	1910000	100	143	4	3	2	2	2024	2200	55282	Maguwoharjo, Sleman	https://www.rumah123.com/properti/yogyakarta/h...
2	Dijual Rumah Baru Mewah Maguwoharjo Include Ko...	1940000	150	110	3	2	2	2	2024	2200	55282	Maguwoharjo, Sleman	https://www.rumah123.com/properti/yogyakarta/h...
3	Rumah Mewah Private Pool di Pusat Kota Jogja	1220000	81	120	3	2	2	2	2024	2200	55282	Maguwoharjo, Sleman	https://www.rumah123.com/properti/yogyakarta/h...
4	RUMAH MEWAH KONTEMPORER 2 LANTAI FULLY FURNISH...	1490000	133	125	4	3	2	2	2024	3500	55282	Maguwoharjo, Sleman	https://www.rumah123.com/properti/sleman/hos17...

Gambar. 4. Data Setelah Preprocesing

Selanjutnya dataset dibagi menjadi dua subset: data pelatihan (training set) dan data pengujian (testing set). Menggunakan, 70-90% data digunakan untuk pelatihan, sementara 10-30% sisanya digunakan untuk pengujian.

D. *Multiple Linear Regression*

Regresi linier berganda adalah bentuk yang lebih kompleks dari model linier sederhana [16]. Regresi Linear Berganda (MLR) adalah salah satu metode statistik yang digunakan untuk mengembangkan model prediksi dengan melibatkan lebih dari satu variabel independen. Dalam teknik ini, model prediksi dibentuk dengan mempertimbangkan pengaruh simultan dari berbagai variabel independen yang terlibat. Dengan menggunakan pendekatan

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 10, No. 4, December 2025, Pp. 3301-3312

regresi linear berganda, prediksi tidak hanya bergantung pada satu variabel, tetapi pada berbagai variabel independen yang dianalisis secara bersamaan [17]. Hal ini memungkinkan analisis yang lebih mendalam dan menyeluruh mengenai hubungan antara satu variabel dependen (terikat) dengan sejumlah variabel independen(bebas. Metode ini menggunakan persamaan matematis untuk menggambarkan hubungan antara variabel dependen (Y) dengan beberapa variabel independen (X1, X2, ..., Xn), di mana persamaan regresi linear berganda ini dapat dituliskan pada persamaan (1):

$$Y = \alpha + \beta 1\, X1 + \beta 2\, X2 + \beta n\, Xn + e \qquad (1)$$

Penjelasan:
Y = Variabel dependen (terikat) β = Slope atau nilai koefisien regresi
X = Variabel independen (bebas) e = Kesalahan (Error)
α = Konstanta (Intercept)

Dengan pemilihan model *MLR* memungkinkan analisis hubungan linear antara variabel dependen (harga rumah) dan beberapa variabel independen seperti dijelaskan pada tabel 1. Koefisien dalam MLR memiliki interpretasi yang jelas, yaitu perubahan rata-rata dalam variabel dependen untuk setiap unit perubahan dalam variabel independen, dengan asumsi variabel lain konstan. Hal ini memungkinkan pemahaman yang lebih mendalam mengenai faktor-faktor yang paling mempengaruhi harga properti. MLR mengasumsikan hubungan linear antara variabel dependen dan independen. Dalam dunia nyata, hubungan ini mungkin tidak selalu linear, sehingga dapat menyebabkan kurang akuratnya prediksi. MLR mungkin tidak cukup efektif dalam menangkap pola yang lebih kompleks dalam data, terutama ketika data memiliki interaksi yang rumit atau struktur non-linear.

E. Random Forest Regressor

Random forest adalah sebuah model ansambel yang terdiri dari sejumlah pohon keputusan yang digunakan dalam metode regresi dan klasifikasi. Model ini mengimplementasikan teknik bootstrap aggregating (bagging) serta pemilihan fitur secara acak untuk meningkatkan akurasi prediksi [18]. Random forest memiliki sejumlah keunggulan, seperti kemampuannya mendeteksi kesalahan yang cukup besar, memberikan kinerja klasifikasi yang optimal, dapat menangani data dengan jumlah sampel yang terbatas, serta efektif dalam mengestimasi data yang hilang [19]. Pohon Keputusan mengelompokkan sampel data yang kelasnya tidak diketahui ke dalam kelas yang sudah ada. Tujuan dari Pohon Keputusan adalah untuk menghindari overfitting pada dataset dan mencapai akurasi tertinggi. Adapun persamaan pada model random forest ini pada persamaan (2) :

$$\hat{y}i = \frac{1}{N_{tree\,e}} \sum_{n=1}^{N_{tree}} \hat{Y}n \qquad (2)$$

Dengan : $\hat{y}i$ = Hasil prediksi, N_tree= Total nilai pohon, $\hat{Y}n$ = Hasil prediksi pohon ke-n.

Dengan pemilihan algoritma ini dapat menangkap hubungan yang kompleks dan non-linear dalam data, menjadikannya sangat efektif untuk prediksi harga rumah yang dipengaruhi oleh berbagai faktor yang tidak selalu memiliki hubungan linear. Dengan menggabungkan hasil dari banyak pohon keputusan, RFR mengurangi risiko overfitting, terutama pada data dengan banyak variabel dan fitur. Pada algoritma RFR, performa model sangat dipengaruhi oleh pemilihan hyperparameter. Beberapa hyperparameter utama yang perlu disetel diantaranya meliputi jumlah pohon (n_estimators), kedalaman maksimal pohon (max_depth), jumlah fitur yang dipertimbangkan untuk split di setiap node (max_features). Untuk mendapatkan performa optimal, RFR membutuhkan penyesuaian hyperparameter yang lebih rumit, seperti n_estimators dan kedalaman max_depth. Hasil dari banyak pohon keputusan sulit diterjemahkan menjadi wawasan yang mudah dipahami mengenai pengaruh masing-masing variabel. Adapun langkah-langkah yang dapat digunakan untuk menentukan nilai optimal dari hyperparameter ter-sebut seperti menggunakan teknik Grid Search yaitu teknik brute-force yang mencoba semua kombinasi hyperparameter yang telah ditentukan dalam ruang pencarian yang terbatas [20]. Misalnya, untuk menentukan jumlah pohon (n_estimators) dan kedalaman maksimal pohon (max_depth), kita dapat mendefinisikan be-berapa nilai potensial dan Grid Search akan menguji setiap kombinasi untuk mencari hasil terbaik.

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 10, No. 4, December 2025, Pp. 3301-3312

III. HASIL DAN PEMBAHASAN

A. Dataset Scraping

Setelah data melalui tahapan mulai dari pengumpulan data dengan menggunakan teknik web scraping terhadap website penjualan properti rumah yang ada diyogyakarta pada website *www.rumah123.com*. Data tersebut kemudian melalui tahapan pra-pemrosesan. Langkah ini bertujuan untuk meningkatkan kualitas dan keakuratan data sebelum digunakan dalam proses analisis atau pemodelan lebih lanjut. Dataset yang dihasilkan dari tahapan tersebut berformat Comma Separated Values (CSV), yang terdiri dari 506 baris dan 10 kolom. Fitur dan variabel dalam dataset ini meliputi harga, luas lahan, luas bangunan, jumlah kamar tidur, jumlah kamar mandi, carport, jumlah lantai, tahun, daya listrik, dan kode pos. Gambaran lengkap mengenai fitur dan variabel, serta isi dari dataset yang telah diperoleh, dapat dilihat pada gambar 5.

index	harga	lahan	bangunan	kamar_tidur	kamar_mandi	carport	Lantai	tahun	listrik	kode_pos
0	1850000	118	126	4	3	1	2	2024	4400	55282
1	1910000	100	143	4	3	2	2	2024	2200	55282
2	1940000	150	110	3	2	2	2	2024	2200	55282
3	1220000	81	120	3	2	2	2	2024	2200	55282
4	1490000	133	125	4	3	2	2	2024	3500	55282

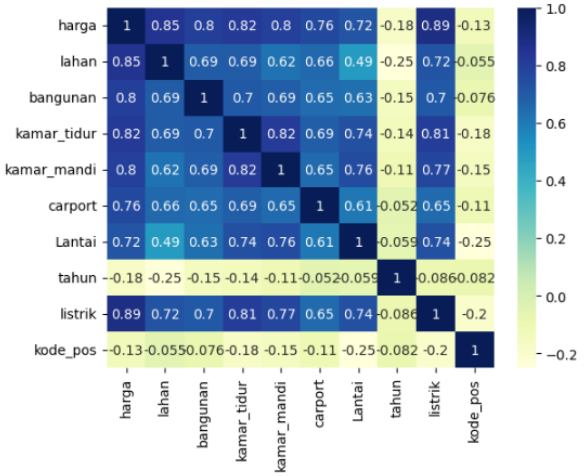
```
(506, 10)
```

Gambar. 5. Dataset rumah jogja

Dalam analisis data, anomali atau outlier adalah data atau observasi yang menyimpang secara signifikan dari pola umum dalam dataset. Jika ada rumah dengan harga yang jauh lebih tinggi atau rendah dibandingkan dengan rumah lain dalam dataset, ini dapat diidentifikasi sebagai anomali. Pemeriksaan lebih lanjut mungkin diperlukan untuk melihat apakah ini merupakan kesalahan input (misalnya, salah pengetikan) atau memang properti tersebut memiliki karakteristik unik yang membuatnya sangat mahal atau murah. Jika ada properti dengan fitur seperti luas lahan atau bangunan yang tidak masuk akal (misalnya, 10 kali lipat lebih besar atau lebih kecil dari rata-rata), ini juga dapat dianggap sebagai anomali. Penanganan anomali data dengan benar dapat membantu meningkatkan keakuratan model dan memastikan bahwa prediksi yang dihasilkan lebih dapat diandalkan.

B. Matriks Korelasi

Merupakan tabel yang menampilkan hubungan linier antara dua atau lebih variabel dalam sebuah dataset, menggunakan koefisien korelasi Pearson yang berkisar dari -1 hingga +1. Koefisien positif menunjukkan adanya hubungan langsung antara variabel, di mana peningkatan satu variabel diikuti oleh peningkatan variabel lainnya. Sebaliknya, koefisien negatif menunjukkan hubungan terbalik, di mana kenaikan satu variabel menyebabkan penurunan pada variabel lain. Nilai 0 mengindikasikan tidak adanya hubungan linier antara variabel-variabel tersebut.



Gambar. 6. Matriks Korelasi Variabel

Pada gambar 6 menunjukan nilai serta korelasi atau kedekatan nilai antar setiap variabel. Adapaun Variabel dengan korelasi tertinggi menunjukkan hubungan positif sempurna di mana peningkatan satu variabel diikuti oleh peningkatan variabel lainnya terhadap variabel harga rumah diantaranya variabel listrik dengan nilai 0.89, lahan 0.85, kamar_tidur 0.82, bangunan 0.8 dan variabel kamar_mandi 0.8 . Serta Koefisien -1 menunjukkan hubungan negatif sempurna, di mana peningkatan satu variabel menyebabkan penurunan variabel lain seperti variabel

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 10, No. 4, December 2025, Pp. 3301-3312

kode_pos dengan nilai -0.13 dan variabel tahun dengan nilai -0.18.

C. Dataset Split

Dataset ini terdiri dari 506 baris yang berisi informasi terkait berbagai properti rumah, dengan total 10 kolom yang masing-masing mewakili fitur atau atribut yang penting dalam penilaian harga rumah. Dalam rangka membangun dan mengevaluasi model prediksi, dataset ini dibagi menjadi dua bagian yaitu data latih dan data uji. Tujuan utamanya adalah untuk memastikan bahwa model dapat dilatih dengan cukup data (untuk belajar dari pola) dan diuji dengan data yang tidak terlihat (untuk mengevaluasi generalisasi). Adapun dilakukan 3 skenario pembagian dataset diantaranya pembagian 70:30: Pembagian ini adalah yang paling umum digunakan, Ini memberikan keseimbangan yang baik antara ukuran set pelatihan dan set pengujian. 80:20: Digunakan ketika dataset cukup besar dan model membutuhkan lebih banyak data untuk dilatih. Pembagian ini memastikan bahwa model memiliki lebih banyak data untuk belajar, tetapi tetap memberikan cukup data untuk pengujian. 90:10: Digunakan ketika data sangat terbatas, dan penting untuk memberikan sebanyak mungkin data ke dalam set pelatihan. Meskipun demikian, set pengujian yang lebih kecil dapat membuat evaluasi model menjadi kurang stabil atau kurang representatif. Pembagian ini dilakukan untuk memastikan bahwa model yang dihasilkan dapat melakukan generalisasi dengan baik pada data baru yang belum pernah dianalisis sebelumnya, sehingga mampu memberikan prediksi yang akurat. Adapun pembagian data training dan data testing dapat dilihat pada tabel 2.

TABEL II
DATASET SPLIT

Dataset	Training		Testing	
	Persen %	Data	Persen %	Data
506	70	354	30	152
	80	405	20	101
	90	455	10	51

D. Pengujian menggunakan *Multiple Linear Regression*

Dalam proses pengujian dataset menggunakan algoritma *Multiple Linear Regression*, dilakukan pemanggilan fungsi Linear Regression yang diakses melalui library sklearn pada bahasa pemrograman Python. Proses ini melibatkan tahapan yang mencakup import library yang diperlukan, kemudian memanfaatkan fungsi tersebut untuk membangun model prediksi berdasarkan data yang telah dibagi sebelumnya. Source code yang digunakan dalam pengujian ini dapat dilihat pada Gambar 7, yang mengilustrasikan bagaimana algoritma ini diterapkan secara praktis untuk melakukan regresi linear pada dataset rumah.

```
model = LinearRegression()
model.fit(x_train, y_train)
```

Gambar. 7. Model *Multiple Linear Regression*

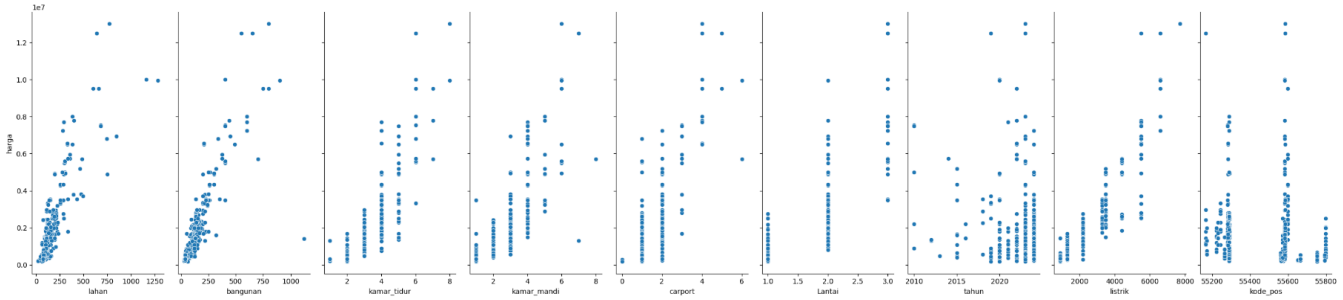
Setelah Pemodelan selesai dilanjutkan tahap evaluasi terhadap model *Multiple Linear Regression (MLR)* diantaranya seperti uji Koefisiensi Determinasi atau sering disebut sebagai *R-squared* (R^2), Tahapan untuk mengukur pengaruh antara dua variabel yang membantu menilai akurasi algoritma linear regression. Koefisien ini memiliki nilai antara 0 hingga 1. Jika $R^2 = 0$, berarti tidak ada hubungan antara variabel independen dan dependen, sedangkan jika $R^2 = 1$ atau mendekati 1, menunjukkan hubungan yang semakin kuat antara kedua variabel tersebut. Evaluasi model juga menggunakan Metrik Evaluasi *Mean Square Error* (MAE) memberikan informasi tentang rata-rata kesalahan yang diharapkan, sedangkan *Root Mean Square Error* (RMSE) lebih fokus pada kesalahan yang lebih besar, sehingga lebih sensitif terhadap data yang tidak normal atau outliers.

TABEL III
MATRIK EVALUASI MULTIPLE LINEAR REGRESSION

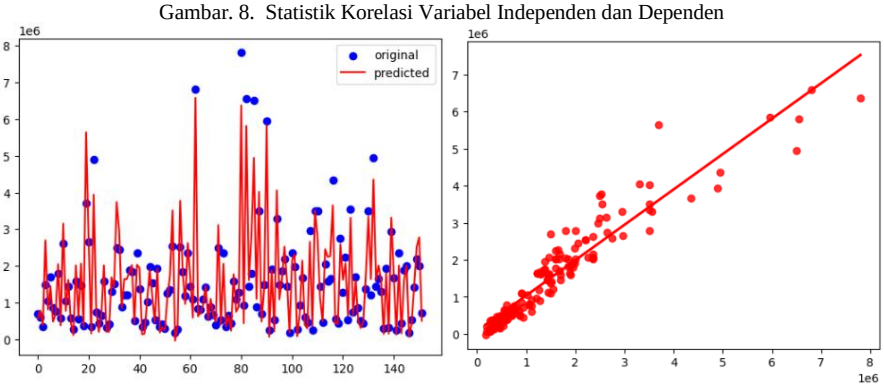
Model	Data Split	Random state	Waktu Komputasi	Model Evaluasi		
				MAE	RMSE	R-Squared (R^2)
Multiple Linear Regression	90:10	100	0,5``	288476.77	457563.99	0.757061
	80:20	100	0,4``	308761.48	472343.62	0.894612
	70:30	100	0,4``	296048.93	430296.63	0.900158

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 10, No. 4, December 2025, Pp. 3301-3312

Berdasarkan hasil pada tabel 3, didapatkan model *MLR* dengan kofisien Determinasi (R^2) terbesar menggunakan percobaan dengan data split 70% training dan 30% Testing, menghasilkan tingkat akurasi prediksi sebesar 0.900158 atau 90%, dengan nilai RMSE sebesar 430296.63 atau Rp. 430.296,63 dan nilai MAE sebesar 296048.93. Adapun visualisasi korelasi variabel terhadap variabel dependen (harga) beserta diagram statis yang menunjukkan titik ketelitian antara nilai asli dengan nilai prediksi dapat dilihat pada gambar 8 & 9.



Gambar. 8. Statistik Korelasi Variabel Independen dan Dependen



Gambar. 9. Statistik data Aktual dan Indeks Prediksi

Multiple Linear Regression (MLR) sangat ideal untuk data yang memiliki hubungan linear yang jelas antara variabel-variabelnya, serta memberikan kemudahan dalam interpretasi hasil karena model ini menghasilkan koefisien langsung untuk setiap variabel. Namun, MLR memiliki keterbatasan dalam hal fleksibilitas, terutama ketika menghadapi data dengan hubungan non-linear, yang membuatnya kurang efektif dalam memodelkan data semacam itu. Selain itu, MLR cenderung rentan terhadap pengaruh outlier, yang dapat mengganggu hasil analisis, serta rentan terhadap masalah multikolinearitas, di mana variabel independen memiliki korelasi tinggi satu sama lain, yang dapat mengakibatkan ketidakstabilan dalam estimasi model.

E. Pengujian menggunakan *Random Forest Regressor*

Pengujian dataset dengan menggunakan algoritma *Random Forest Regressor* dilakukan dengan memanfaatkan fungsi bawaan dari algoritma *Random Forest* yang diimplementasikan melalui *library sklearn* pada bahasa pemrograman Python. Proses ini melibatkan pemanggilan fungsi-fungsi dalam sklearn untuk membangun dan melatih model Random Forest, yang kemudian digunakan untuk memprediksi nilai berdasarkan data yang diberikan. Dengan menambah jumlah jumlah pohon (n_estimators), kedalaman maksimal pohon (max_depth), jumlah fitur yang dipertimbangkan untuk split di setiap node (max_features), model menjadi lebih kompleks dan mampu menangkap lebih banyak variasi dalam data. Namun, peningkatan jumlah pohon juga meningkatkan waktu komputasi dan memori yang dibutuhkan. Source code implementasi dapat dilihat pada gambar 10.

```
# Membuat pipeline dengan scaling dan Random Forest
pipeline = Pipeline([
    ('scaler', StandardScaler()),
    ('model', RandomForestRegressor(n_estimators=100, max_depth=30,
                                    max_features=0.5, max_samples=0.5,
                                    random_state=100))
])
```

```
Pipeline
Pipeline(steps=[('standardscaler', StandardScaler()),
                ('randomforestregressor',
                 RandomForestRegressor(max_depth=30, max_features=0.5,
                                       max_samples=0.5, random_state=100))])
StandardScaler
RandomForestRegressor
```

Gambar. 10. Model Random Forest Regressor

Pipeline memungkinkan untuk menggabungkan beberapa langkah, seperti prapemrosesan dan model machine learning "*Random Forest Regressor*" dalam satu objek. Ini mempermudah proses pelatihan dan prediksi. Untuk

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 10, No. 4, December 2025, Pp. 3301-3312

StandardScaler memastikan bahwa semua fitur memiliki skala yang sama, yang sering kali penting dalam algoritma machine learning tertentu yang sensitif terhadap skala data. Selanjutnya memasuki tahapan untuk menentukan nilai optimal dari hyperparameter dengan menggunakan teknik *Grid Search*, adapun source code dan hasil dengan teknik *Grid Search* pada algoritma *Random Forest Regressor* dapat dilihat pada gambar 11.

```
from sklearn.model_selection import GridSearchCV
param_grid = {
    'randomforestregressor__max_depth': [10, 15, 20,30,35,40,45,50],
    'randomforestregressor__n_estimators': [20, 30, 40,50,70,80,100],
}
grid_search = GridSearchCV(rf_model, param_grid, cv=5, scoring='neg_mean_squared_error')
grid_search.fit(X_train, y_train)
print("Best parameters:", grid_search.best_params_)
```

```
Best parameters: {'randomforestregressor__max_depth': 10, 'randomforestregressor__n_estimators': 100}
```

Gambar. 11. Source Code Grid Search RFR

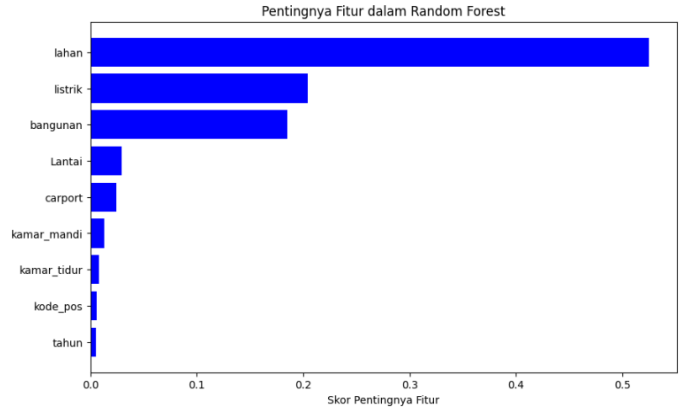
Setelah Pemodelan selesai dilanjutkan tahap evaluasi terhadap model *Random Forest Regressor (RFR)*, Seperti pada Model *MLR*, Evaluasi pada Model *RFR* juga menggunakan uji Koefisiensi Determinasi atau R-squared (R^2), menggunakan Metrik Evaluasi *Mean Square Error* (MAE) dan *Root Mean Square Error* (RMSE). Hasil Model Evaluasi dapat dilihat pada tabel 4.

TABEL IV
MATRIK EVALUASI RANDOM FOREST REGRESSOR

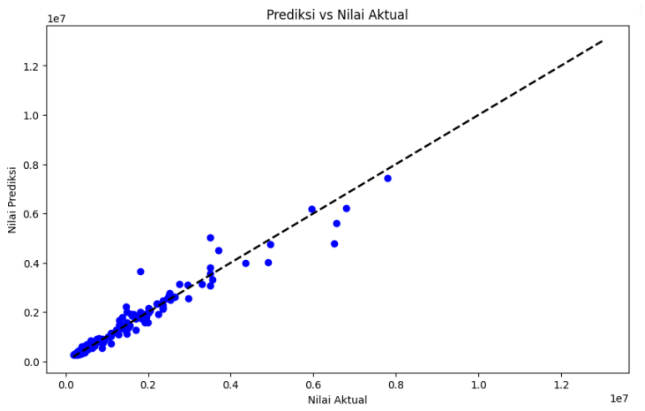
Model	Data Split	N estimators	Max Depth	Waktu Komputasi	Model Evaluasi		
					MAE	RMSE	R-Squared (R^2)
Random Forest Regressor	90:10	40	20	29``	155742.20	273116.06	0.913446
	80:20	40	20	28``	189986.87	355700.40	0.940235
	70:30	100	10	28``	175840.89	325115.44	0.943003

Berdasarkan hasil pada tabel 4, didapatkan model *RFR* dengan kofisien Determinasi (R^2) terbesar menggunakan percobaan dengan data split 70% training dan 30% Testing, dengan N Estimator = 100 menghasilkan tingkat akurasi prediksi sebesar 0.943003 atau 94,3%, dengan nilai RMSE sebesar 325115.44 atau Rp. 325.115,44 dan nilai MAE sebesar 175840.89 dengan waktu komputasi 28``(Detik). Model ini membutuhkan lebih banyak waktu dan sumber daya komputasi, terutama untuk dataset besar dengan banyak pohon, membuat proses pelatihan dan prediksi lebih lambat.

Adapun diagram visualisasi Pentingnya Fitur atau variabel independen terhadap variabel dependen (harga) dapat dilihat pada gambar 12, dimana fitur yang paling berpengaruh terhadap harga rumah dengan menggunakan Model *Random Forest Regressor (RFR)* terdapat pada Fitur Lahan(Luas Lahan). Selanjutnya diagram statis yang menunjukkan titik antara nilai aktual dengan nilai prediksi yang mendekati presisi linear dapat dilihat pada gambar 13.



Gambar. 12. Pentingnya Fitur dalam Random Forest.



Gambar.13. Prediksi vs Nilai Aktual

Random Forest lebih fleksibel dan dapat menangani data dengan hubungan non-linear, tetapi memiliki risiko overfitting, sulit diinterpretasikan, dan membutuhkan lebih banyak sumber daya komputasi.

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 10, No. 4, December 2025, Pp. 3301-3312

F. Evaluasi Perbandingan Algoritma

Adapaun hasil evaluasi perbandingan Model Algortima *Multiple Linear Regression* dan *Random Forest Regressor* yang telah digunakan terhadap Estimasi harga Rumah dengan teknik *Web Scraping* terhadap data website penjualan rumah dikawasan yogyakarta mendapatkan hasil seperti pada pada tabel 5 .

TABEL V
MATRIK EVALUASI *MLR* DAN *RFR*

Model	Random State	Waktu Komputasi	Model Evaluasi		
			MAE	RMSE	R-Squared (R^2)
Multiple Linear Regression	100	0,4``	296048.93	430296.63	0.900158
Random Forest Regressor	100	28``	175840.89	325115.44	0.943003

Hasil dari evaluasi pada tabel 5 menunjukkan bahwa Algoritma *Random Forest Regressor* memberikan performa yang lebih baik dalam mengestimasi harga properti rumah. Algoritma ini berhasil mencapai Mean Absolute Error (MAE) sebesar 175840.89, Root Mean Square Error (RMSE) sebesar 325115.44, dan R-Squared sebesar 0.943003. Disisi lain, model Multiple Linear Regression tampil dengan performa yang sedikit kurang optimal, ditunjukkan oleh MAE sebesar 296048.93, RMSE sebesar 430296.63, dan R-Squared sebesar 0.900158. Hal ini menandakan bahwa Multiple Linear Regression memiliki tingkat kesalahan prediksi yang sedikit lebih tinggi dan kemampuan yang lebih terbatas dalam menjelaskan variabilitas harga properti rumah dibandingkan dengan *Random Forest Regressor*.

Dengan hasil yang diperoleh pada penelitian ini dapat memberikan hasil prediksi yang baik bagi praktisi properti, calon pembeli, dan pembuat kebijakan untuk membuat keputusan yang lebih baik, berdasarkan prediksi harga properti rumah yang lebih akurat. Bagi praktisi, ini berarti penetapan harga dan pemasaran yang lebih efektif. Bagi calon pembeli, ini berarti keputusan pembelian yang lebih cerdas dan perencanaan keuangan yang lebih matang. Sementara itu, bagi pembuat kebijakan, penelitian ini menawarkan wawasan untuk merumuskan kebijakan perumahan yang lebih tepat sasaran dan responsif terhadap dinamika pasar. Penelitian ini memberikan kontribusi dengan menyajikan analisis komprehensif tentang performa algoritma *MLR* dan *Random Forest Regressor* dalam memprediksi harga properti di Yogyakarta, yang memiliki dinamika pasar yang khas. Berbeda dari penelitian terdahulu [7][8][9] yang lebih banyak berfokus pada suatu area yang berbeda, menggunakan dataset serta fitur dengan karakteristik yang berbeda, serta hasil model evaluasi yang mendapatkan peningkatan hasil. Penelitian ini menawarkan perspektif baru melalui penerapan teknik web scraping pada data real-time dari Website di Yogyakarta dan menguji kinerja model tersebut pada data ini.

IV. KESIMPULAN

Model Random Forest Regressor menunjukkan performa yang lebih unggul dibandingkan dengan Multiple Linear Regression dalam konteks estimasi harga properti rumah. Keunggulan ini terlihat dari perbandingan nilai evaluasi yang digunakan, yaitu Mean Absolute Error (MAE), Root Mean Square Error (RMSE), dan R-Squared (R^2), di mana Random Forest Regressor memiliki nilai yang lebih baik. Secara spesifik, Random Forest Regressor menghasilkan MAE sebesar 175840.89, RMSE sebesar 325115.44, dan R-Squared sebesar 0.943003. Nilai-nilai ini menunjukkan bahwa model tersebut mampu memprediksi harga properti dengan tingkat akurasi yang lebih tinggi walau memerlukan waktu komputasi yang lebih lama dengan kecepatan komputasi 28``. Sebaliknya, model Multiple Linear Regression menunjukkan performa yang sedikit kurang optimal dengan MAE sebesar 296048.93, RMSE sebesar 430296.63, dan R-Squared sebesar 0.900158. Oleh karena itu, dapat disimpulkan bahwa Random Forest Regressor lebih efektif dan dapat diandalkan untuk estimasi harga properti rumah di Yogyakarta, terutama berdasarkan data yang diperoleh melalui teknik web scraping dari situs web penjualan rumah, dalam konteks penggunaan data hasil web scraping dari website penjualan rumah.. Penggunaan Random Forest Regressor dalam skenario ini memberikan hasil yang lebih akurat, menjadikannya pilihan yang lebih tepat dalam analisis harga properti.

DAFTAR PUSTAKA

[1] Saiful, A., Andryana, S., & Gunaryati, A. (2021). Prediksi harga rumah menggunakan web scrapping dan machine learning dengan algoritma linear regression. JATISI (Jurnal Teknik Informatika dan Sistem Informasi), 8(1), 41-50.
[2] Suudiah, Vina Apriliani.. (2023). Analisis Faktor-Faktor Yang Mempengaruhi Harga Rumah Di Kota Batam. Tractare: Jurnal Ekonomi-Manajemen, 5(2), 119-138.
[3] Ridho, Ihda Innar, Galih Mahalisa, Dwi Retno Sari, and Ihsanul Fikri. (2022). Metode Neural Network Untuk Penentuan Akurasi Prediksi Harga Rumah. Technologia: Jurnal Ilmiah, 13(1), 56-58.
[4] Arlintang, Nadifa Oksa, Lucia Rita Indrawati, and Yustirania Septiani.. (2020). Analisis Pengaruh Investasi, Belanja Modal, dan Infrastruktur Ekonomi Terhadap Pertumbuhan Ekonomi di Daerah Istimewa Yogyakarta Tahun 2008-2018. DINAMIC: Directory Journal of Economic, 2(2), 573-586.
[5] Lawson, R. (2015). Web Scraping With Python: Scrape Data from Any Website With the Power of Phython, Packt Publishing Ltd, Birmingham UK
[6] Yuantari, C., & Handayani, S. (2017). Textbook of Descriptive & Inferential Statistics.
[7] Haryanto, Cep, Nining Rahaningsih, and Fadhil Muhammad Basysyar (2023). Komparasi Algoritma Machine Learning Dalam Memprediksi Harga Rumah. JATI (Jurnal Mahasiswa Teknik Informatika), 7(1), 533-539.
[8] Özalp, A. Y., & Akıncı, H. (2024). Comparison of tree-based machine learning algorithms in price prediction of residential real estate. Gümüşhane Üniversitesi Fen Bilimleri Dergisi, 14(1), 116-130.
[9] Abdul-Rahman, S., Zulkifley, N. H., Ismail, I., & Mutalib, S. (2021). Advanced machine learning algorithms for house price prediction: Case study in Kuala Lumpur. International Journal of Advanced Computer Science and Applications, 12(12).
[10] Ho, Winky KO, Bo-Sin Tang, and Siu Wai Wong. (2021). Predicting property prices with machine learning algorithms. Journal of Property Research, 38(1), 48-70.
[11] Sirisuriya, De S. (2015). A comparative study on web scraping.
[12] Hafiz, Y. A., and Endah Sudarmilah. (2023). Implementasi Web Scraping Pada Portal Berita Online. Inisiasi, 55-60.
[13] Tandra, Dave Julianno, Agustinus Noertjahyana, and Anita Nathania Purbowo. (2020). Implementasi Web Scraping untuk Pengumpulan Informasi Promo Makanan Menggunakan Klasifikasi Naïve Bayes. Jurnal Infra, 8(1), 289-294.
[14] Sobirin, Rahmat, Dimas Prayoga, Muhammad Abdul Basit, and Kusrini Kusrini.. (2023, November). Forecasting the effect of parameters on AQI values with Machine learning: Multiple Linear Regression. In 2023 6th International Conference on Information and Communications Technology (ICOIACT) (pp. 159-164). IEEE.
[15] García, Salvador, Sergio Ramírez-Gallego, Julián Luengo, José Manuel Benítez, and Francisco Herrera. (2016). Big data preprocessing: methods and prospects. Big data analytics, 1, 1-22.
[16] Puteri, K., & Silvanie, A. (2020). Machine learning untuk model prediksi harga sembako dengan metode regresi linear berganda. Jurnal Nasional Informatika (JUNIF), 1(2), 82-94.
[17] Ayuni, Ghebyla Najla, and Devi Fitrianah. (2019). Penerapan metode Regresi Linear untuk prediksi penjualan properti pada PT XYZ. Jurnal telematika, 14(2), 79-86.
[18] Purwa, Taly. (2019). Perbandingan Metode Regresi Logistik dan Random Forest untuk Klasifikasi Data Imbalanced (Studi Kasus: Klasifikasi Rumah Tangga Miskin di Kabupaten Karangasem, Bali Tahun 2017). Jurnal Matematika, Statistika dan Komputasi, 16(1), 58-73.
[19] Ali, J., Khan, R., Ahmad, N., & Maqsood, I. (2012). Random forests and decision trees. International Journal of Computer Science Issues (IJCSI), 9(5), 272.
[20] Han, Sunwoo, and Hyunjoong Kim. (2021). Optimal feature set size in random forest regression. *Applied Sciences*, *11*(8), 3428.