ISSN: 2598-8565 (media cetak)
ISSN: 2620-4339 (media online)

PREDIKSI HARGA RESALE SEPATU ADIDAS SAMBA OG MENGGUNAKAN METODE RANDOM FOREST REGRESSION BERDASARKAN VARIABEL E-COMMERCE SEBAGAI REFERENSI HARGA UNTUK SELLER DAN BUYER

Milhan Muid✉, Fida Maisa Hana, Widya Cholid Wahyudin
Ilmu Komputer, Fakultas Sains dan Teknologi, Universitas Muhammadiyah Kudus, Kudus, Indonesia
Email: milhanmuid@gmail.com

DOI: https://doi.org/10.46880/jmika.Vol10No2.pp770-780

ABSTRACT

E-commerce has expanded the fashion market, including Adidas Samba OG resale activities, where prices vary across listings. Such variation can make it difficult for buyers to assess price fairness and for sellers to determine selling prices. This study aims to estimate Adidas Samba OG resale prices using Random Forest Regression based on Size, Color, E-commerce Platform, and Units Sold. The dataset comprises 999 listings manually collected in June 2026: 399 from Shopee, 240 from Tokopedia, and 360 from TikTok Shop. The analysis included preprocessing, an 80:20 data split, baseline modeling, hyperparameter optimization through GridSearchCV, model evaluation, feature importance analysis, visualization, and new-data prediction. The baseline model produced an MAE of IDR 29,884.25, RMSE of IDR 111,979.93, and R^2 of 0.8401. After tuning, the model yielded an MAE of IDR 32,907.91, RMSE of IDR 104,414.54, and R^2 of 0.8610. Units Sold showed the highest contribution at 58.99%. Predictions for five scenarios ranged from IDR 2,138,312 to IDR 2,404,340. The results indicate that the model can provide an additional reference for buyers and sellers when considering Adidas Samba OG resale prices.

Keywords: Random Forest Regression, Resale Price Prediction, Sneaker, E-Commerce, Adidas Samba OG.

ABSTRAK

Perdagangan berbasis e-commerce turut memperluas pasar produk fashion, termasuk aktivitas resale Adidas Samba OG yang memiliki perbedaan harga antar-listing. Variasi tersebut dapat menyulitkan buyer dalam menilai kewajaran harga serta seller dalam menetapkan harga jual. Penelitian ini dilakukan untuk mengestimasi harga resale Adidas Samba OG dengan menerapkan Random Forest Regression menggunakan variabel Size, Warna, Platform e-commerce, dan Jumlah_Terjual. Dataset penelitian terdiri atas 999 listing yang dikumpulkan secara manual pada Juni 2026, meliputi 399 data dari Shopee, 240 dari Tokopedia, dan 360 dari TikTok Shop. Tahapan analisis mencakup preprocessing, pembagian data 80:20, pemodelan baseline, optimasi hyperparameter melalui GridSearchCV, evaluasi, analisis feature importance, visualisasi, serta prediksi data baru. Model baseline memperoleh MAE Rp29.884,25, RMSE Rp111.979,93, dan R^2 0,8401. Setelah proses tuning, diperoleh MAE Rp32.907,91, RMSE Rp104.414,54, dan R^2 0,8610. Jumlah_Terjual memiliki kontribusi tertinggi sebesar 58,99%. Penerapan pada lima skenario menghasilkan estimasi harga Rp2.138.312–Rp2.404.340. Hasil tersebut menunjukkan bahwa model dapat dimanfaatkan sebagai referensi tambahan bagi buyer dan seller dalam mempertimbangkan harga resale Adidas Samba OG.

Kata Kunci: Random Forest Regression, Prediksi Harga Resale, Sneaker, E-Commerce, Adidas Samba OG.

PENDAHULUAN

Perkembangan teknologi digital telah mendorong perubahan pola perdagangan melalui pemanfaatan *e-commerce*, yang memberikan kemudahan bagi konsumen untuk memperoleh dan membandingkan produk tanpa terbatas wilayah. Kondisi tersebut didukung oleh tingginya penggunaan internet di Indonesia. (APJI, 2024) mencatat jumlah pengguna internet pada 2024 mencapai 221.563.479 jiwa dengan tingkat penetrasi sebesar 79,5%, yang menunjukkan besarnya aktivitas masyarakat dalam ekosistem digital (APJI, 2024; Laudon & Traver, 2024). Perkembangan ini turut memperluas pasar *fashion*, khususnya *sneaker*, yang tidak hanya memiliki fungsi penggunaan tetapi juga nilai gaya hidup dan ekonomi. Salah satu bentuk aktivitas perdagangan yang berkembang adalah pasar *resale*, termasuk Adidas Samba OG. Penelitian (Raditya et al., 2021) menunjukkan bahwa data penjualan historis dapat dimanfaatkan untuk memprediksi harga *resale sneaker*, dengan *Random Forest* memberikan hasil yang lebih baik dibandingkan regresi linear.

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Adidas Samba OG memiliki aktivitas perdagangan pada berbagai platform *e-commerce*, sehingga harga *resale* dapat berbeda antar-*seller* dan platform. Faktor seperti *size*, warna, jumlah terjual, persepsi nilai produk, serta karakteristik pemasaran dapat berkaitan dengan keputusan dan minat pembelian produk Adidas (Kusumawati & Hasan, 2025; Pangestu & Erpurini, 2026; Polgan et al., 2025). Perbedaan harga tersebut dapat menyulitkan *buyer* dalam menilai kewajaran harga dan *seller* dalam menentukan harga jual yang sesuai. Oleh karena itu, informasi yang terdapat pada *listing e-commerce*, seperti *Size*, Warna, Platform, dan Jumlah_Terjual, dapat dimanfaatkan sebagai dasar untuk membangun estimasi harga yang lebih sistematis. Pendekatan berbasis data diperlukan agar proses penentuan harga tidak hanya bergantung pada pengamatan subjektif, tetapi dapat mempertimbangkan pola yang terbentuk dari data historis.

Machine learning memungkinkan komputer mempelajari pola dan hubungan antarvariabel dari data untuk menghasilkan prediksi serta mendukung pengambilan keputusan (Jordan & Mitchell, 2015). Penelitian ini menggunakan *Random Forest Regression*, yaitu metode *ensemble* yang menggabungkan beberapa *decision tree* untuk menghasilkan prediksi melalui proses agregasi. Random Forest memiliki kemampuan dalam menangani hubungan *non-linear* dan memanfaatkan pengacakan data maupun fitur untuk menghasilkan model yang lebih stabil (Breiman, 2001). Implementasinya juga tersedia dalam pustaka *scikit-learn* sehingga dapat digunakan dalam pengembangan model prediksi berbasis data (Pedregosa et al., 2011).

Sejumlah penelitian telah memanfaatkan *Random Forest* untuk mengestimasi harga pada berbagai objek, di antaranya rumah, laptop, dan properti (Lestari & Astuti, 2022; Hadi & Benedict, 2024; Gede et al., 2025; Hayya et al., 2025; Septiyanah, 2025; Tian, 2024; Balqis et al., 2026; Putri Utami et al., 2025). Pemanfaatan metode tersebut juga dijumpai pada bidang *e-commerce*, seperti pengelolaan persediaan dan klasifikasi performa penjualan produk (Ridwan, 2024; Nisa et al., 2026). Dalam konteks *resale sneaker*, Raditya et al. (2021) menunjukkan bahwa informasi penjualan pada periode sebelumnya dapat digunakan untuk memperkirakan harga *resale*, dengan *Random Forest* menunjukkan kinerja yang lebih baik dibandingkan regresi linear.

Berdasarkan kajian terdahulu yang telah dibahas, objek penelitian, sumber data, dan konteks prediksi yang digunakan masih berbeda dengan fokus penelitian ini. Kajian yang secara khusus mengestimasi harga *resale* Adidas Samba OG menggunakan *Random Forest Regression* dengan memanfaatkan atribut *listing* dari beberapa *platform e-commerce* Indonesia masih terbatas. Pembahasan mengenai peran variabel *Size*, Warna, Platform, dan Jumlah_Terjual dalam estimasi harga *resale* Adidas Samba OG juga belum banyak ditemukan pada penelitian yang dikaji. Kondisi tersebut menunjukkan adanya celah penelitian yang menjadi dasar pelaksanaan penelitian ini.

Pemilihan Adidas Samba OG sebagai objek penelitian didasarkan pada keberadaan produk tersebut di berbagai *platform e-commerce* dengan harga *resale* yang dapat bervariasi antar-*listing*. Selain itu, beberapa faktor seperti *size*, warna, jumlah terjual, persepsi terhadap nilai produk, serta karakteristik pemasaran diketahui berkaitan dengan keputusan maupun minat pembelian produk Adidas (Kusumawati & Hasan, 2025; Pangestu & Erpurini, 2026; Polgan et al., 2025). Data dari Shopee, Tokopedia, dan TikTok Shop digunakan agar penelitian tidak hanya bergantung pada satu sumber serta dapat mencakup perbedaan karakteristik *listing* antarplatform. Kebaruan penelitian ini terletak pada penerapan *Random Forest Regression* untuk mengestimasi harga *resale* Adidas Samba OG dengan memanfaatkan gabungan data dari tiga *platform e-commerce* Indonesia serta variabel *Size*, Warna, Platform, dan Jumlah_Terjual sebagai prediktor. Kontribusi setiap variabel dianalisis melalui *feature importance*, sedangkan kemampuan prediksi model diukur menggunakan MAE, RMSE, dan R^2 serta dilengkapi dengan visualisasi dan penerapan pada data baru. Berdasarkan pendekatan tersebut, penelitian ini bertujuan menghasilkan model estimasi harga *resale* Adidas Samba OG yang dapat dimanfaatkan sebagai referensi pendukung bagi *buyer* dan *seller* dalam mempertimbangkan harga.

TINJAUAN PUSTAKA

Artificial Intelligence dan *Machine Learning*

Artificial Intelligence (AI) merupakan bidang ilmu komputer yang memungkinkan sistem mengenali pola, menganalisis data, dan menghasilkan prediksi secara otomatis. Penerapannya pada *e-commerce* mencakup rekomendasi produk, analisis perilaku konsumen, dan prediksi harga (Géron, 2019; Han et al., 2012; Kotu & Deshpande, 2015; Laudon & Traver, 2024; Russell & Norvig, 2021). Salah satu bagian AI adalah *Machine Learning*, yang memungkinkan komputer mempelajari pola dari data historis untuk menghasilkan prediksi. Pendekatan *supervised learning* digunakan ketika model mempelajari hubungan antara variabel masukan dan target yang telah diketahui (Achmad Ridwan et al., 2025; Géron,

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

ISSN: 2598-8565 (media cetak)
ISSN: 2620-4339 (media online)

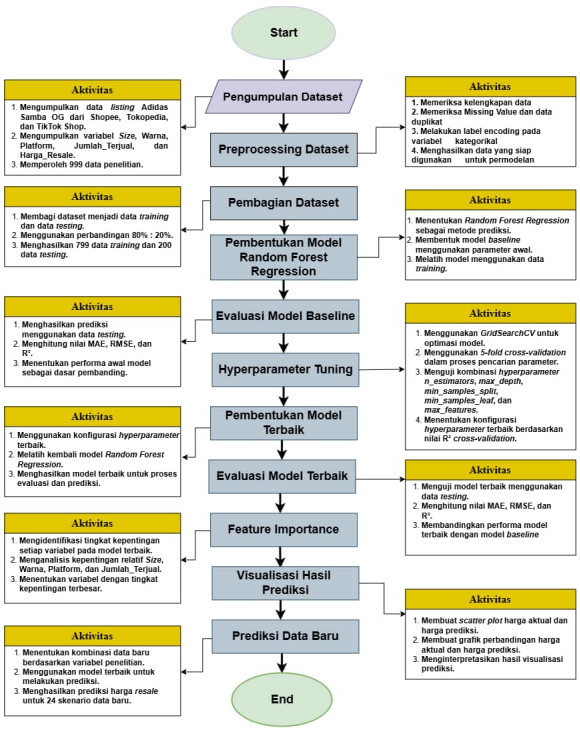
2019; Kumar, 2014; Russell & Norvig, 2021). Penelitian ini menerapkan *supervised learning* untuk memprediksi harga *resale* Adidas Samba OG.

Random Forest Regression

Random Forest Regression merupakan metode *ensemble learning* yang membangun beberapa *decision tree* dan menggabungkan hasil prediksinya untuk memperoleh estimasi yang lebih stabil (Breiman, 2001; Géron, 2019; Kumar, 2014). Metode ini mampu menangani hubungan *non-linear* serta variabel numerik dan kategorikal (Gede et al., 2025; Kotu & Deshpande, 2015; Wahyudin et al., 2025). Implementasinya dapat dilakukan menggunakan *scikit-learn* (Pedregosa et al., 2011). Penerapan Random Forest juga menunjukkan hasil yang baik pada klasifikasi penyakit jantung dengan akurasi 90,7% (Wahyudin et al., 2025). Dalam penelitian ini, Random Forest Regression digunakan untuk memprediksi harga *resale* berdasarkan *size*, warna, *platform e-commerce*, dan jumlah terjual (Gede et al., 2025; Han et al., 2012; Septiyanah, 2025; Wahyudin, 2026).

METODE PENELITIAN

Penelitian ini dilaksanakan melalui beberapa tahapan yang tersusun secara sistematis, yaitu pengumpulan data, *preprocessing data*, pelatihan model dengan metode *Random Forest Regression*, evaluasi model, serta prediksi dan analisis hasil. Alur penelitian secara keseluruhan dapat dilihat pada Gambar 1.



Gambar 1. Flowchart Alur Penelitian

Berdasarkan Gambar 1, alur penelitian yang telah digambarkan secara umum tersebut selanjutnya diuraikan lebih rinci pada masing-masing tahapan berikut, mulai dari proses pengumpulan data hingga tahap prediksi dan analisis hasil.

Pengumpulan Data

Pengumpulan data dilakukan secara manual pada Juni 2026 melalui tiga *platform e-commerce* Indonesia, yaitu Shopee, Tokopedia, dan TikTok Shop. Data yang digunakan terdiri atas 999 *listing* Adidas Samba OG dalam kondisi baru (BNIB), dengan rincian 399 data berasal dari Shopee, 240 data dari Tokopedia, dan 360 data dari TikTok Shop. Atribut yang dicatat mencakup *Size*, Warna, Platform *e-commerce*, Jumlah_Terjual, serta Harga_Resale. Kesesuaian produk dan kondisi barang ditentukan berdasarkan informasi yang tersedia pada judul dan deskripsi setiap *listing*. Seluruh data selanjutnya melalui tahap pemeriksaan untuk memastikan kelengkapan dan kesesuaiannya dengan kriteria penelitian. Data yang tidak lengkap, tercatat lebih dari satu kali, atau tidak memenuhi kriteria penelitian dikeluarkan sebelum proses *preprocessing*. Data yang lolos proses seleksi kemudian disimpan dalam format *.csv* dan digunakan pada tahapan analisis berikutnya.

Penelitian ini menggunakan dua jenis variabel, yaitu variabel independen (X) sebagai input model dan variabel dependen (Y) sebagai target prediksi, seperti pada Tabel 1.

Tabel 1. Variabel Penelitian

No	Variabel	Simbol	Jenis
1	*Size*	X1	Independen
2	Warna	X2	Independen
3	Platform *E-Commerce*	X3	Independen
4	Jumlah Terjual	X4	Independen
5	Harga *Resale*	Y	Dependen

Berdasarkan Tabel 1, variabel *Size*, Warna, Platform *e-commerce*, dan Jumlah_Terjual digunakan sebagai variabel independen, sedangkan Harga_Resale digunakan sebagai variabel dependen yang menjadi target prediksi model *Random Forest Regression*. Setelah variabel penelitian ditetapkan, data selanjutnya diproses pada tahap *preprocessing* sebelum digunakan dalam pemodelan.

Preprocessing Data

Tahap *preprocessing* dilakukan untuk memastikan data telah berada dalam kondisi yang sesuai sebelum digunakan pada proses pemodelan

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Random Forest Regression. Langkah awal mencakup pemeriksaan terhadap kemungkinan adanya *missing value* pada seluruh variabel yang digunakan. Setelah itu, data diperiksa untuk menemukan pencatatan yang berulang atau duplikat, kemudian atribut yang tidak berperan dalam proses prediksi, seperti kolom nomor urut atau kolom tambahan, dihapus dari dataset. Selanjutnya, variabel kategorikal berupa Warna dan Platform *e-commerce* diubah ke dalam bentuk numerik melalui teknik *label encoding* agar dapat diproses oleh algoritma. Setelah seluruh proses selesai, struktur dan tipe data kembali diperiksa untuk memastikan dataset telah siap digunakan pada tahap pembagian data dan pembangunan model.

Pembagian Dataset

Dataset yang telah selesai melalui tahap *preprocessing* selanjutnya dipisahkan menjadi dua bagian, yaitu data *training* dan data *testing*, dengan menggunakan metode *train-test split*. Proporsi pembagian ditetapkan sebesar 80:20, sehingga dari 999 data diperoleh 799 data untuk proses pelatihan dan 200 data untuk proses pengujian. Pembagian data dilakukan dengan menetapkan random_state = 42 agar hasil pemisahan tetap konsisten ketika proses dijalankan kembali. Data *training* digunakan untuk membentuk pola pembelajaran pada model *Random Forest Regression*, sedangkan data *testing* dimanfaatkan untuk menilai kemampuan model dalam menghasilkan prediksi terhadap data yang tidak dilibatkan pada tahap pelatihan.

Pembentukan Model *Random Forest Regression*

Setelah proses pembagian dataset selesai, model *Random Forest Regression* dibangun menggunakan data *training*. Variabel *Size*, Warna, Platform *e-commerce*, dan Jumlah_Terjual digunakan sebagai variabel independen, sedangkan Harga_Resale ditetapkan sebagai variabel target. Pada proses pemodelan, *Random Forest Regression* membentuk sejumlah *decision tree* berdasarkan data pelatihan, kemudian menggabungkan keluaran prediksi dari seluruh pohon untuk memperoleh nilai estimasi akhir. Mekanisme tersebut dinyatakan melalui persamaan berikut:

$$\hat{y} = \frac{1}{k}\sum_{i=1}^{k} T_i(x)$$

Keterangan:

$\hat{y}$: nilai prediksi akhir harga *resale* Adidas Samba OG
k : jumlah *decision tree* dalam model
$T_i(x)$: hasil prediksi dari *decision tree* ke-i
x : data masukan berupa size (X1), warna (X2), platform *e-commerce* (X3), dan jumlah terjual (X4)

Nilai prediksi akhir (ŷ) diperoleh dari hasil agregasi prediksi seluruh *decision tree* yang terbentuk selama proses pelatihan. Model dilatih menggunakan X_train dan y_train untuk mempelajari hubungan antara variabel prediktor dan Harga_Resale. Setelah tahap pelatihan selesai, model diterapkan pada data *testing* (X_test) untuk menghasilkan nilai prediksi (y_pred). Nilai tersebut kemudian dibandingkan dengan harga aktual (y_test) dan digunakan sebagai dasar pengukuran kinerja model melalui MAE, RMSE, dan R^2.

Evaluasi *Model Baseline*

Evaluasi terhadap model *baseline* dilakukan menggunakan data *testing* untuk mengetahui kemampuan awal *Random Forest Regression* dalam memperkirakan harga *resale* Adidas Samba OG sebelum proses *hyperparameter tuning*. Hasil prediksi model (y_pred) dibandingkan dengan nilai aktual (y_test) menggunakan tiga metrik, yaitu *Mean Absolute Error* (MAE), *Root Mean Square Error* (RMSE), dan koefisien determinasi (R^2). MAE digunakan untuk menghitung rata-rata selisih absolut antara nilai prediksi dan nilai sebenarnya, sedangkan RMSE digunakan untuk mengukur besarnya kesalahan prediksi dengan memberikan bobot lebih besar pada selisih yang bernilai tinggi. Adapun R^2 digunakan untuk menilai sejauh mana model mampu menjelaskan variasi Harga_Resale pada data pengujian.

1. *Mean Absolute Error* (MAE)

Mean Absolute Error (MAE) digunakan untuk mengukur rata-rata selisih absolut antara nilai aktual dan nilai prediksi tanpa memperhatikan arah kesalahan. Semakin kecil nilai MAE, semakin baik kemampuan model dalam memprediksi harga *resale* Adidas Samba OG.

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i|$$

n : Jumlah data pengujian (*testing* data).
Σ : Simbol penjumlahan seluruh data pengujian
i : Indeks data ke-i.
y_i : Nilai aktual harga *resale* pada data ke-i.
$\hat{y}_i$: Nilai prediksi harga *resale* yang dihasilkan oleh model pada data ke-i.
$|y_i - \hat{y}_i|$: Nilai absolut dari selisih antara nilai aktual dan nilai hasil prediksi.

2. *Root Mean Square Error* (RMSE)

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Root Mean Square Error (RMSE) digunakan untuk mengukur tingkat kesalahan prediksi model dengan memberikan bobot lebih besar pada kesalahan bernilai tinggi. Semakin kecil nilai RMSE, semakin baik performa model *Random Forest Regression* dalam memprediksi harga *resale* Adidas Samba OG.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_{i-}\hat{y}_i)^2}$$

Keterangan:
n : Jumlah data pengujian (*testing data*).
Σ : Simbol penjumlahan seluruh data pengujian.
i : Indeks data ke i
y_i : Nilai aktual harga *resale* pada data ke-i.
$\hat{y}_i$: Nilai prediksi harga *resale* yang dihasilkan oleh model pada data ke-i.
$(y_{i-}\hat{y}_i)$ ²: Kuadrat selisih antara nilai aktual dan nilai hasil prediksi.

3. Koefisien Determinasi (R²)

Koefisien Determinasi (R²) digunakan untuk mengukur kemampuan model dalam menjelaskan hubungan antara variabel independen dan variabel dependen. Nilai R² berada pada rentang 0 hingga 1, di mana nilai yang semakin mendekati 1 menunjukkan model semakin baik dalam menjelaskan variasi data. Dalam penelitian ini, R² digunakan untuk mengetahui seberapa baik model *Random Forest Regression* menjelaskan pengaruh variabel *size*, warna, platform *e-commerce*, dan jumlah terjual terhadap harga *resale* Adidas Samba OG.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

Keterangan:
n : Jumlah data pengujian (*testing* data).
Σ : Simbol penjumlahan seluruh data pengujian.
i : Indeks data ke-i
y_i : Nilai aktual harga *resale* pada data ke-i.
$\bar{y}$: Nilai rata-rata seluruh data aktual harga resale.
$\hat{y}_i$: Nilai prediksi harga *resale* yang dihasilkan oleh model pada data ke-i.
$(y_i - \bar{y})$: Jumlah kuadrat selisih antara nilai aktual dan rata-rata nilai aktual (Total Sum of Squares).
$(y_i - \hat{y}_i)$: Jumlah kuadrat selisih antara nilai aktual dan nilai hasil prediksi (Residual Sum of Squares).

Hasil evaluasi model *baseline* kemudian digunakan sebagai acuan untuk membandingkan perubahan kinerja model setelah dilakukan *hyperparameter tuning*.

Hyperparameter Tuning

Penentuan konfigurasi parameter dilakukan menggunakan *GridSearchCV* pada model *Random Forest Regression*. Proses pencarian menerapkan skema *5-fold cross-validation* dengan R² sebagai dasar penilaian. Ruang pencarian parameter mencakup n_estimators = [100, 200, 300, 500], max_depth = [10, 15, 20, 25, None], min_samples_split = [2, 3, 5, 8, 10], min_samples_leaf = [1, 2, 3, 4], serta max_features = [1.0, sqrt, log2]. Untuk menjaga konsistensi hasil pemodelan, estimator menggunakan random_state = 42. Kombinasi seluruh parameter menghasilkan 1.200 konfigurasi yang masing-masing dievaluasi pada lima *fold*, sehingga total proses yang dijalankan mencapai 6.000 *fitting*. Setiap konfigurasi dibandingkan berdasarkan nilai rata-rata R² dari hasil *cross-validation*, kemudian kombinasi dengan nilai tertinggi dipilih sebagai konfigurasi terbaik untuk digunakan pada tahap pembentukan model berikutnya.

Pembentukan Model Terbaik

Setelah konfigurasi *hyperparameter* terbaik diperoleh dari proses *GridSearchCV*, konfigurasi tersebut diterapkan untuk membangun ulang model *Random Forest Regression*. Proses pelatihan dilakukan menggunakan data *training* yang sama, yaitu X_train sebagai data masukan dan y_train sebagai target. Model yang telah selesai dilatih kemudian digunakan untuk menghasilkan prediksi terhadap X_test. Data *testing* yang digunakan tetap mengacu pada pembagian data sebelumnya agar perbandingan performa antara model *baseline* dan model hasil *hyperparameter tuning* dilakukan pada kondisi pengujian yang konsisten. Model yang terbentuk selanjutnya digunakan dalam evaluasi kinerja, analisis *feature importance*, visualisasi hasil prediksi, serta pengujian pada data baru.

Evaluasi Model Terbaik

Model *Random Forest Regression* yang diperoleh setelah proses *hyperparameter tuning* selanjutnya diuji menggunakan data *testing* yang sama seperti pada evaluasi model *baseline*. Hasil prediksi terhadap X_test dibandingkan dengan nilai aktual (y_test) menggunakan tiga ukuran kinerja, yaitu *Mean Absolute Error* (MAE), *Root Mean Square Error* (RMSE), dan koefisien determinasi (R²). Nilai evaluasi tersebut kemudian dibandingkan dengan hasil model *baseline* untuk mengidentifikasi perubahan kinerja setelah proses optimasi parameter.

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

MAE dan RMSE digunakan untuk melihat besarnya kesalahan prediksi, di mana nilai yang lebih kecil menunjukkan tingkat kesalahan yang lebih rendah. Sementara itu, R^2 digunakan untuk menilai kemampuan model dalam menjelaskan variasi Harga_Resale pada data pengujian, dengan nilai yang lebih tinggi menunjukkan kemampuan penjelasan yang lebih besar. Model yang telah dievaluasi selanjutnya digunakan pada analisis *feature importance*, visualisasi hasil prediksi, serta penerapan pada data baru.

Feature Importance

Feature importance digunakan untuk mengetahui tingkat kepentingan relatif setiap variabel (*Size*, Warna, *Platform*, dan Jumlah_Terjual) terhadap prediksi model *Random Forest Regression* terbaik, di mana variabel dengan nilai *feature importance* lebih tinggi menunjukkan pengaruh yang lebih besar dalam pembentukan prediksi. Hasil analisis ini divisualisasikan dalam bentuk grafik untuk mempermudah interpretasi dan digunakan sebagai informasi tambahan dalam memahami karakteristik model saat memprediksi harga *resale* Adidas Samba OG.

Visualisasi Hasil Prediksi

Hasil prediksi dari model terbaik pada data *testing* divisualisasikan untuk membandingkan nilai Harga_Resale aktual (y_test) dengan nilai yang dihasilkan model (y_pred). Penyajian dilakukan melalui *scatter plot* untuk menggambarkan hubungan antara harga aktual dan harga prediksi, serta grafik perbandingan untuk memperlihatkan perbedaan keduanya pada setiap data pengujian. Pada *scatter plot*, posisi titik yang semakin dekat dengan garis acuan menunjukkan tingkat kesesuaian yang semakin tinggi antara nilai aktual dan hasil prediksi. Sementara itu, grafik perbandingan digunakan untuk meninjau variasi serta selisih nilai pada masing-masing data uji. Hasil visualisasi kemudian digunakan sebagai informasi pendukung dalam menafsirkan performa model bersama metrik MAE, RMSE, dan R^2 sebelum model diterapkan pada proses prediksi data baru.

Prediksi Data Baru

Setelah model terbaik selesai dievaluasi, model tersebut digunakan untuk memperkirakan Harga_Resale pada lima skenario input yang telah ditentukan. Setiap skenario dibentuk dari kombinasi variabel *Size*, Warna, Platform *e-commerce*, dan Jumlah_Terjual. Untuk menjaga kesesuaian dengan data yang digunakan saat pelatihan, variabel kategorikal pada skenario baru dikonversi menggunakan skema *encoding* yang sama seperti pada tahap *preprocessing*.

Selanjutnya, kelima kombinasi input tersebut diproses oleh model untuk menghasilkan estimasi Harga_Resale. Skenario yang digunakan berfungsi untuk menunjukkan penerapan model pada variasi input baru dan tidak berasal dari *listing* baru yang dikumpulkan langsung dari *marketplace*. Nilai prediksi yang dihasilkan kemudian disajikan dan dibahas pada bagian Hasil dan Pembahasan.

HASIL DAN PEMBAHASAN

Hasil Pengumpulan Data

Berdasarkan proses pengumpulan data, diperoleh sebanyak 999 *listing* Adidas Samba OG dari Shopee, Tokopedia, dan TikTok Shop pada periode Juni 2026. Data terdiri atas 399 data dari Shopee, 240 data dari Tokopedia, dan 360 data dari TikTok Shop. Data yang digunakan memenuhi kriteria berupa produk Adidas Samba OG dalam kondisi baru (*BNIB*) serta memiliki informasi *Size*, Warna, Jumlah_Terjual, dan Harga_Resale. Identifikasi produk dan kondisi dilakukan berdasarkan informasi pada judul dan deskripsi *listing*. Data duplikat dan *listing* yang tidak memenuhi kriteria penelitian dikeluarkan sebelum tahap *preprocessing*, sehingga diperoleh 999 data yang digunakan dalam penelitian, dengan cuplikan sebagian data ditunjukkan pada Tabel 2.

Tabel 2. Dataset Adidas Samba OG

Size	Warna	Platform	Jumlah Terjual	Harga Resale
37.33	Black White	Shopee	1000	2240000
38	Black White	Shopee	1000	2630000
38.67	Black White	Shopee	1000	2500000
39.33	Black White	Shopee	1000	2630000
40	Black White	Shopee	1000	1910000

***Preprocessing* Data**

Tahap *preprocessing* dilakukan untuk menyiapkan dataset agar memiliki kondisi dan susunan data yang sesuai sebelum masuk ke proses pemodelan. Proses ini diawali dengan *data cleaning* untuk menyeleksi serta menghapus data yang tidak lengkap atau tidak memenuhi kriteria penelitian. Selanjutnya, dilakukan pemeriksaan terhadap keberadaan *missing value* dan data duplikat. Atribut yang tidak memiliki peran dalam proses pemodelan kemudian dihilangkan sehingga dataset hanya menyisakan variabel yang relevan dengan kebutuhan penelitian.

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Tahap berikutnya dilakukan transformasi terhadap variabel kategorikal, yaitu Warna dan Platform *e-commerce*, ke dalam representasi numerik dengan menggunakan teknik *label encoding*. Konversi ini diperlukan agar kedua variabel tersebut dapat dikenali dan diolah oleh algoritma *Random Forest Regression*. Berdasarkan hasil pemeriksaan, dataset tidak ditemukan mengandung *missing value* maupun data duplikat. Seluruh kategori pada variabel Warna dan Platform juga telah berhasil diubah ke bentuk numerik, dengan hasil pemetaan *label encoding* ditampilkan pada Tabel 3.

Tabel 3. Hasil Label Encoding Variabel Kategorikal

Variabel	Kategori	Nilai Encoding
Warna	Black White	0
Warna	Navy White	1
Warna	White Black	2
Warna	White Green	3
Platform	Shopee	0
Platform	TikTok Shop	1
Platform	Tokopedia	2

Berdasarkan hasil *preprocessing*, dataset telah memiliki struktur dan tipe data yang sesuai untuk digunakan pada tahap pembagian dataset dan pemodelan *Random Forest Regression*.

Pembagian Dataset

Setelah proses *preprocessing* selesai, data dipisahkan ke dalam dua kelompok untuk keperluan pelatihan dan pengujian model. Pemisahan dilakukan dengan pendekatan *train-test split* menggunakan proporsi 80% untuk data *training* dan 20% untuk data *testing*, serta menetapkan *random state* sebesar 42 agar pembagian data dapat direproduksi secara konsisten. Dari total 999 data, sebanyak 799 data digunakan dalam proses pembelajaran model, sedangkan 200 data lainnya dialokasikan untuk tahap pengujian. Data *training* dimanfaatkan untuk membentuk pola hubungan antara variabel prediktor dengan Harga_Resale, sementara data *testing* digunakan untuk menilai kinerja model pada data yang tidak dilibatkan selama proses pelatihan.

Pembentukan Model *Random Forest Regression*

Setelah pembagian *dataset*, model *Random Forest Regression* dibangun dan dilatih (*fitting*) menggunakan data *training*, dengan *X_train* (Size, Warna, Platform, Jumlah_Terjual) sebagai variabel independen dan *y_train* (Harga_*Resale*) sebagai variabel target. Variabel Jumlah_Terjual digunakan sebagai salah satu bagian dari data masukan, sedangkan y_train yang berisi Harga_Resale berperan sebagai target yang dipelajari oleh model. Pada tahap pelatihan, *Random Forest Regression* membentuk sekumpulan *decision tree* dari data *training* untuk menangkap pola keterkaitan antara variabel prediktor dan Harga_Resale.

Setelah proses pembelajaran selesai, model diterapkan pada X_test, yaitu data pengujian yang tidak digunakan pada tahap pelatihan. Nilai keluaran yang dihasilkan dari proses tersebut kemudian disimpan dalam y_pred sebagai hasil prediksi model. Nilai *y_pred* kemudian dibandingkan dengan nilai aktual (*y_test*) pada tahap evaluasi untuk menghitung metrik MAE, RMSE, dan R^2.

Evaluasi Model Baseline

Evaluasi model *baseline* dilakukan menggunakan MAE, RMSE, dan R^2 untuk mengukur kinerja awal *Random Forest Regression* pada data *testing*.

Tabel 4. Hasil Evaluasi Model Baseline

Model	MAE	RMSE	R^2
Random Forest Regression Baseline	Rp29.884,25	Rp111.979,93	0,8401

Pengujian terhadap model *baseline* menghasilkan MAE sebesar Rp29.884,25, RMSE sebesar Rp111.979,93, dan R^2 sebesar 0,8401. Nilai MAE tersebut menunjukkan bahwa secara rata-rata, selisih absolut antara Harga_Resale yang diprediksi dengan harga sebenarnya berada di kisaran Rp29.884,25. Adapun RMSE sebesar Rp111.979,93 memperlihatkan bahwa masih terdapat sejumlah prediksi dengan tingkat kesalahan yang lebih besar, mengingat metrik ini lebih sensitif terhadap selisih prediksi yang bernilai tinggi. Sementara itu, nilai R^2 sebesar 0,8401 menunjukkan bahwa sekitar 84,01% variasi Harga_Resale pada data pengujian dapat dijelaskan oleh model, sedangkan sekitar 15,99% sisanya belum dapat diterangkan oleh pola yang dipelajari. Hasil tersebut digunakan sebagai acuan performa awal *Random Forest Regression* sebelum dilakukan *hyperparameter tuning*, sehingga perubahan kinerja setelah proses optimasi dapat dibandingkan secara lebih jelas.

Hyperparameter Tuning

Tahap *hyperparameter tuning* dilakukan menggunakan *GridSearchCV* dengan menerapkan *5-fold cross-validation* dan R^2 sebagai kriteria evaluasi. Proses pencarian melibatkan lima *hyperparameter*, yaitu n_estimators, max_depth, min_samples_split, min_samples_leaf, dan max_features. Seluruh nilai

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

yang ditetapkan dalam ruang pencarian membentuk 1.200 kombinasi konfigurasi. Karena setiap kombinasi diuji pada lima *fold*, keseluruhan proses menghasilkan 6.000 kali *fitting*. Pemilihan konfigurasi dilakukan berdasarkan nilai rata-rata R^2 tertinggi yang diperoleh selama proses *cross-validation*. Parameter dengan hasil terbaik selanjutnya ditampilkan pada Tabel 5.

Tabel 5. Hasil Hyperparameter Tuning

No	Hyperparameter	Nilai Terbaik
1	max_depth	20
2	max_features	1,0
3	min_samples_leaf	1
4	min_samples_split	8
5	n_estimators	100

Berdasarkan hasil pada Tabel 5, konfigurasi dengan nilai evaluasi terbaik terdiri atas max_depth = 20, max_features = 1.0, min_samples_leaf = 1, min_samples_split = 8, dan n_estimators = 100, dengan Best CV R^2 sebesar 0,7781. Konfigurasi tersebut membentuk model menggunakan 100 *decision tree* dengan batas kedalaman maksimum sebanyak 20 tingkat. Nilai min_samples_split = 8 menunjukkan bahwa suatu *node* memerlukan sedikitnya delapan sampel sebelum dapat dilakukan pemisahan, sedangkan min_samples_leaf = 1 memungkinkan terbentuknya *leaf node* dengan minimal satu sampel. Pengaturan max_features = 1.0 memungkinkan seluruh fitur yang tersedia dipertimbangkan ketika menentukan pemisahan pada setiap *node*. Sementara itu, Best CV R^2 sebesar 0,7781 merupakan nilai rata-rata R^2 yang diperoleh selama proses validasi silang dan digunakan untuk menentukan konfigurasi terbaik, sehingga nilai tersebut tidak dianggap sebagai hasil akhir evaluasi pada data *testing*. Konfigurasi yang terpilih kemudian diterapkan untuk membangun kembali model *Random Forest Regression* sebelum dilakukan pengujian terhadap data *testing*.

Pembentukan dan Prediksi Menggunakan Model Terbaik

Konfigurasi *hyperparameter* yang memperoleh hasil terbaik dari proses *GridSearchCV*, yaitu max_depth = 20, max_features = 1.0, min_samples_leaf = 1, min_samples_split = 8, dan n_estimators = 100, kemudian digunakan untuk membangun kembali model *Random Forest Regression*. Proses pelatihan dilakukan menggunakan data *training* yang sama seperti pada model *baseline* agar perbedaan performa yang muncul dapat diamati pada kondisi data yang konsisten. Setelah model selesai dilatih, proses prediksi dilakukan terhadap data *testing* yang sebelumnya tidak digunakan selama tahap pelatihan. Data pengujian yang digunakan juga sama dengan data pada evaluasi model *baseline*, sehingga hasil kedua model dapat dibandingkan secara setara. Nilai prediksi yang diperoleh selanjutnya digunakan sebagai dasar perhitungan MAE, RMSE, dan R^2 untuk menilai kinerja model setelah penerapan *hyperparameter tuning*.

Evaluasi Model Hasil Hyperparameter Tuning

Evaluasi model terbaik dilakukan menggunakan tiga metrik, yaitu MAE, RMSE, dan R^2.

Tabel 6. Hasil Evaluasi Setelah Hyperparameter Tuning

Model	MAE	RMSE	R^2
Random Forest Regression setelah Hyperparameter Tuning	Rp32.907,91	Rp104.414,54	0,8610

Hasil pengujian menunjukkan bahwa penerapan *hyperparameter tuning* memberikan perubahan yang tidak sama pada seluruh metrik evaluasi. Nilai RMSE berkurang dari Rp111.979,93 pada model *baseline* menjadi Rp104.414,54, sementara nilai R^2 meningkat dari 0,8401 menjadi 0,8610. Peningkatan R^2 tersebut menunjukkan bahwa model setelah proses *tuning* mampu menjelaskan sekitar 86,10% variasi Harga_Resale pada data pengujian, lebih tinggi dibandingkan model *baseline* yang mampu menjelaskan 84,01%. Penurunan RMSE juga memperlihatkan bahwa model hasil *tuning* lebih mampu menekan kesalahan prediksi dengan selisih yang relatif besar. Di sisi lain, nilai MAE justru meningkat dari Rp29.884,25 menjadi Rp32.907,91, yang menunjukkan bahwa rata-rata selisih absolut antara hasil prediksi dan harga aktual menjadi lebih besar. Perbedaan arah perubahan ketiga metrik tersebut dapat dikaitkan dengan proses pemilihan konfigurasi pada *GridSearchCV* yang menggunakan rata-rata R^2 hasil *cross-validation* sebagai dasar optimasi, sehingga pencarian parameter tidak secara khusus diarahkan untuk memperoleh nilai MAE terendah. Dengan demikian, proses *hyperparameter tuning* memberikan perbaikan pada RMSE dan R^2, tetapi belum menghasilkan penurunan pada MAE.

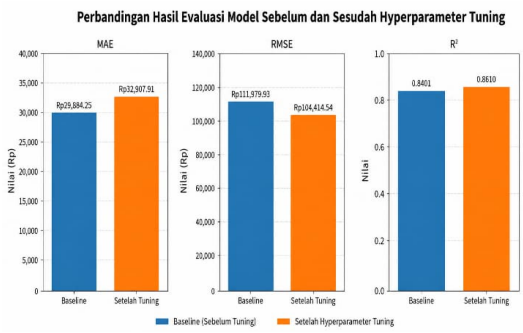
Temuan penelitian ini memiliki keterkaitan dengan Raditya et al. (2021), yang menunjukkan bahwa *Random Forest* dapat digunakan dalam prediksi harga *resale sneaker* dan memberikan hasil yang lebih baik dibandingkan regresi linear. Hasil tersebut juga sejalan dengan Hayya et al. (2025), yang melaporkan bahwa

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

ISSN: 2598-8565 (media cetak)
ISSN: 2620-4339 (media online)

Random Forest memiliki kinerja lebih baik dibandingkan K-Nearest Neighbor (KNN) dalam prediksi harga rumah. Meskipun objek penelitian, karakteristik data, serta variabel yang digunakan berbeda, ketiga penelitian tersebut sama-sama menunjukkan penerapan *Random Forest* pada permasalahan prediksi harga. Dalam penelitian ini, penerapan metode tersebut difokuskan pada estimasi Harga_Resale Adidas Samba OG dengan memanfaatkan atribut *listing* dari Shopee, Tokopedia, dan TikTok Shop.

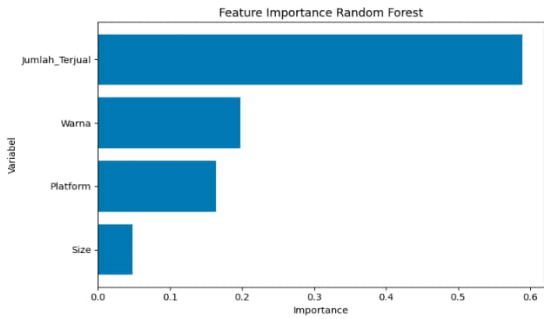
Untuk memberikan gambaran yang lebih jelas mengenai perubahan kinerja model, hasil evaluasi sebelum dan sesudah proses *hyperparameter tuning* divisualisasikan berdasarkan nilai MAE, RMSE, dan R^2. Perbandingan tersebut ditampilkan pada Gambar 2.



Gambar 2. Grafik Perbandingan Evaluasi Model

Analisis *Feature Importance*

Hasil analisis *feature importance* memperlihatkan bahwa setiap variabel memiliki tingkat kontribusi yang berbeda dalam membentuk prediksi Harga_Resale. Variabel Jumlah_Terjual menempati posisi tertinggi dengan nilai kepentingan sebesar 0,589860 atau 58,99%, disusul Warna sebesar 0,197844 atau 19,78%, Platform sebesar 0,164045 atau 16,40%, dan *Size* sebesar 0,048252 atau 4,83%.



Gambar 3. Grafik *Feature Importance*

Temuan tersebut menunjukkan bahwa Jumlah_Terjual merupakan variabel yang paling dominan dalam pola prediksi yang dibangun oleh model *Random Forest Regression*. Besarnya kontribusi variabel ini, yang mencapai lebih dari separuh total nilai kepentingan fitur, menunjukkan bahwa informasi terkait jumlah produk yang telah terjual memiliki peranan relatif lebih besar dibandingkan variabel lainnya. Warna dan Platform tetap memberikan kontribusi terhadap proses prediksi, sedangkan *Size* memiliki tingkat kepentingan paling rendah. Nilai *feature importance* pada penelitian ini menunjukkan besarnya kontribusi relatif setiap variabel terhadap proses pengambilan keputusan model dan tidak dapat diartikan sebagai hubungan sebab-akibat terhadap perubahan Harga_Resale.

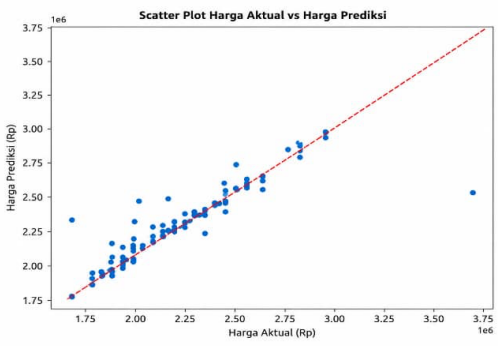
Analisis *feature importance* untuk mengidentifikasi variabel dominan juga digunakan oleh Balqis dan Aini (2026) pada penelitian prediksi harga rumah dengan *Random Forest Regression*. Penelitian tersebut menunjukkan bahwa OverallQual menjadi variabel dengan tingkat kepentingan tertinggi dalam proses prediksi harga rumah. Meskipun objek penelitian dan karakteristik variabel yang digunakan berbeda, kedua penelitian menunjukkan bahwa tingkat kontribusi suatu fitur dalam model *Random Forest Regression* bergantung pada pola dan karakteristik data yang dianalisis. Dalam konteks penelitian ini, dominasi Jumlah_Terjual menunjukkan bahwa informasi aktivitas penjualan memberikan kontribusi paling besar terhadap estimasi Harga_Resale Adidas Samba OG.

Visualisasi Hasil Prediksi

Untuk melihat kesesuaian antara harga aktual dan harga prediksi model pada data pengujian, hasil prediksi divisualisasikan dalam bentuk *scatter plot* serta grafik dan tabel perbandingan.

1. *Scatter Plot* Harga Aktual dan Harga Prediksi

Scatter plot digunakan untuk melihat kedekatan antara harga prediksi dan harga aktual pada setiap data pengujian, di mana semakin dekat titik terhadap garis pembanding, semakin kecil selisih antara nilai prediksi dan nilai aktual.



Gambar 4. *Scatter Plot* Harga Aktual dan Harga Prediksi

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Hasil visualisasi menunjukkan bahwa prediksi model secara umum mengikuti pola harga aktual, meskipun masih terdapat beberapa titik yang menyimpang, menandakan adanya kesalahan prediksi pada sebagian data.

2. Perbandingan Harga Aktual dan Harga Prediksi

Perbedaan nilai error pada setiap data uji dipengaruhi oleh variasi karakteristik *listing*, seperti *Size*, Warna, Platform *e-commerce*, dan Jumlah_Terjual, yang menghasilkan tingkat kesesuaian berbeda terhadap pola yang dipelajari model. Data uji ke-6 menghasilkan prediksi yang sama dengan harga aktual, sedangkan data uji ke-1 memiliki selisih terbesar sebesar Rp60.408. Hasil ini sejalan dengan Raditya et al. (2021) yang menampilkan bahwa *Random Forest* dapat digunakan untuk memprediksi harga *resale sneaker*. Penelitian ini memiliki fokus berbeda dengan menggunakan Adidas Samba OG serta atribut *listing* dari Shopee, Tokopedia, dan TikTok Shop sebagai variabel prediktor.

Tabel 7. Tabel Harga Aktual dan Harga Prediksi

Data Uji	Harga Aktual	Harga Prediksi	Selisih
1	Rp2.549.000	Rp2.488.592	Rp60.408
2	Rp2.275.000	Rp2.268.633	Rp6.367
3	Rp1.799.000	Rp1.807.069	-Rp8.609
4	Rp1.899.000	Rp1.887.591	Rp11.409
5	Rp2.454.000	Rp2.414.649	Rp37.351
6	Rp1.899.000	Rp1.899.000	Rp0

Prediksi Data Baru

Model Random Forest Regression terbaik diterapkan pada lima skenario prediksi yang tidak digunakan dalam proses pelatihan maupun pengujian. Setiap skenario menggunakan kombinasi Size, Warna, Platform *e-commerce*, dan Jumlah_Terjual yang disusun untuk menghasilkan estimasi Harga_Resale. Hasil prediksi menunjukkan perbedaan estimasi harga berdasarkan kombinasi input pada masing-masing skenario. Skenario ini digunakan untuk menunjukkan penerapan model dalam mengestimasi harga *resale* dan tidak dimaksudkan sebagai validasi eksternal terhadap data marketplace.

Tabel 8. Hasil Prediksi Harga 5 Skenario Data Baru

Size	Warna	Platform	Jumlah Terjual	Prediksi Harga
39.33	Black White	Shopee	10	Rp2.138.312
40	White Black	Tokopedia	150	Rp2.191.832
40.67	White Green	Tokopedia	150	Rp2.237.216
41.33	Navy White	TikTok Shop	450	Rp2.292.143
42	Navy White	Shopee	1000	Rp2.404.340

KESIMPULAN

Berdasarkan hasil penelitian, *Random Forest Regression* dapat digunakan untuk memperkirakan Harga_Resale Adidas Samba OG dengan memanfaatkan variabel *Size*, Warna, Platform *e-commerce*, dan Jumlah_Terjual. Model yang diperoleh setelah proses *hyperparameter tuning* menghasilkan MAE sebesar Rp32.907,91, RMSE sebesar Rp104.414,54, dan R^2 sebesar 0,8610. Nilai R^2 tersebut menunjukkan bahwa model mampu menjelaskan sekitar 86,10% variasi Harga_Resale pada data pengujian. Jika dibandingkan dengan model *baseline*, proses *tuning* menghasilkan penurunan RMSE dan peningkatan R^2, meskipun nilai MAE mengalami kenaikan. Hasil analisis *feature importance* menunjukkan bahwa Jumlah_Terjual menjadi variabel dengan kontribusi terbesar sebesar 58,99%, kemudian diikuti Warna sebesar 19,78%, Platform sebesar 16,40%, dan *Size* sebesar 4,83%.

Model yang dihasilkan dapat dimanfaatkan sebagai referensi tambahan bagi *buyer* dalam menilai kewajaran harga serta bagi *seller* dalam mempertimbangkan penetapan harga *resale* Adidas Samba OG. Meskipun demikian, hasil prediksi tetap perlu dipertimbangkan bersama karakteristik *listing* dan kondisi pasar yang berlaku. Penelitian ini masih memiliki keterbatasan karena hanya menggunakan empat variabel prediktor, data dari tiga Platform *e-commerce*, serta berfokus pada satu jenis produk. Selain itu, faktor seperti waktu *listing*, reputasi *seller*, diskon atau promosi, lokasi, dan kondisi pasar belum dimasukkan dalam proses pemodelan. Penelitian selanjutnya dapat memperluas cakupan variabel, sumber data, dan jenis produk serta memasukkan faktor temporal maupun kondisi pasar agar estimasi harga *resale* yang dihasilkan menjadi lebih menyeluruh.

DAFTAR PUSTAKA

Abdullah, M. M., Purwaamijaya, B. M., & Ridlo, M. D. A. (2025). Pengaruh Influencer Marketing Terhadap Purchase Intention Sepatu Adidas Samba di Indonesia. *Jurnal Minfo Polgan, 14*(2), 1665-1671. https://doi.org/10.33395/jmp.v14i2.15144

Achmad, R., Sutikno, T., Riyadi, I., & Cholid Wahyudin, W. (2025). On-Time Student Graduation Prediction Modeling: A Comparative Analysis of Naive Bayes Algorithm and Other Data Mining Classifications. *JOINCS (Journal of Informatics,*

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Network, and Computer Science), *8*(2), 128–135. https://doi.org/10.21070/joincs.v8i2.1679

APJI. (2024). *APJII Jumlah Pengguna Internet Indonesia Tembus 221 Juta Orang*. https://inet.detik.com/cyberlife/

Breiman, L. (2001). *Random Forests* (Vol. 45).

Gede, I. M. P., & Suhartana, I. K. G. (2025). Implementasi Algoritma Random Forest Regression dalam Sistem Prediksi Harga Rumah di Jabodetabek. *Jurnal Nasional Teknologi Informasi dan Aplikasinya, 4*(1), 27-38. https://doi.org/10.24843/JNATIA.2025.v04.i01.p04

Géron, A. (2019). *Hands-on machine learning with scikit-learn, keras, and tensorflow: concepts*. Aurélien Géron-Google Kitaplar.

Han, J., Kamber, M., Pei, J., & Pei, J. (2006). *Data mining: concepts and techniques* (Vol. 2, pp. 383-466). Amsterdam: Elsevier.

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science, 349*(6245), 255-260.

Kotu, V., Deshpande, B., & Deshpande, B. (2014). *Predictive analytics and data mining* (pp. 1-15. Burlington, MA, USA: Morgan Kaufmann.

Kumar, T. S. (2014). *Introduction to Data Mining 1/E*.

Kusumawati, H., & Hasan, W. (2025). Pengaruh Desain Produk, Peran Influencer, Perceived Value, Dan Brand Awareness Terhadap Pembelian Sepatu Adidas. *Jurnal Akuntansi dan Manajemen, 36*(1), 1-13. https://doi.org/10.53916/jam.v36i1.154

Laudon, K. C., & Traver, C. G. (2024). *E-commerce 2023–2024*.

Pangestu, P. B., & Erpurini, W. (2026). The Impact of Electronic Word of Mouth (E-WOM), Brand Image, and Brand Awareness on Consumers' Purchase Intention of Adidas Samba Shoes. *Jurnal Administrasi Bisnis*, *14*(2), 155–164. https://doi.org/10.14710/jab.v14i2.83688

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of machine Learning research, 12*, 2825-2830.

Raditya, D., Nicholas, E. P., Ferarida, A. S., & Hanafiah, N. (2021). Predicting Sneaker Resale Prices using Machine Learning. *Procedia Computer Science*, *179*, 533–540. https://doi.org/10.1016/j.procs.2021.01.037

Russell, S. J., Norvig, P., Canny, J. F., Malik, J. M., & Edwards, D. D. (1995). *Artificial intelligence: a modern approach* (Vol. 2, No. 9). Englewood Cliffs, NJ: Prentice hall.

Septiyanah, S., & Athalina, G. (2025). Implementasi Algoritma Random Forest Regression Dalam Prediksi Harga Laptop. *Jurnal Mnemonic, 8*(1), 70-73. https://doi.org/10.36040/mnemonic.v8i1.12475

Wahyudin, W. C., Sutikno, T., & Umar, R. (2026). A Cluster-Label-Based Framework for Water Quality Risk Pattern Classification Using Naive Bayes and Random Forest. *Jurnal Ilmiah Ilmu Terapan Universitas Jambi, 10*(3), 1511-1522.

Wahyudin, W. C., Sutikno, T., Umar, R., & Ridwan, A. (2025). Comparative Performance Analysis of Data Mining Models for Heart Disease Detection with Feature Selection Implementation. *JOINCS (Journal of Informatics, Network, and Computer Science)*, *8*(1), 87–96.

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.