

Analisis Prediksi Stok Barang Produk Kaos Kaki Berdasarkan Pola Penjualan untuk Optimasi Persediaan Menggunakan Metode Random Forest Regression

Tian Aldi Kurniawan¹ · Hardiansyah²

^{1,2} Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Pamulang, Tangerang Selatan, 15310, Indonesia
e-mail : ¹ tianaldik534@gmail.com , ² dosen01882@unpam.ac.id

Info Artikel

Riwayat Artikel:

Received July 18, 2026

Revised July 29, 2026

Accepted August 20, 2026

Abstract – Improper inventory management can lead to overstock and stockout, disrupting a company's operational activities. PT. Wijaya Konveksi Group faces challenges in determining inventory requirements due to fluctuating demand for its sock products. This study aims to develop a stock prediction model using the Random Forest Regression algorithm based on historical sales data. The data were processed through preprocessing, label encoding, feature engineering, and splitting into training and testing datasets. The model's performance was evaluated using Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and the Coefficient of Determination (R^2). The research results indicate that the hyperparameter tuning process improved model performance, evidenced by a decrease in MAE from 112.5283 to 111.0265, MSE from 26,843.1835 to 25,971.4799, and RMSE from 163.8389 to 161.1567, as well as an increase in the R^2 value from 0.7440 to 0.7524. This coefficient of determination indicates that the model is capable of explaining approximately 75.24% of the variation in stock requirements based on the variables used.

Keywords: Historical Sales Data, Inventory, Machine Learning, Random Forest Regression, Stock Prediction.

Corresponding Author:

Tian Aldi Kurniawan

Email: tianaldik534@gmail.com



This is an open access article under the [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/) license.

Abstrak Indonesia – Pengelolaan persediaan yang kurang tepat dapat menyebabkan overstock maupun stockout sehingga mengganggu operasional perusahaan. PT. Wijaya Konveksi Group menghadapi tantangan dalam menentukan kebutuhan stok karena permintaan produk kaos kaki yang bersifat fluktuatif. Penelitian ini bertujuan membangun model prediksi stok menggunakan algoritma Random Forest Regression berdasarkan data historis penjualan. Data diproses melalui tahapan preprocessing, label encoding, feature engineering, serta pembagian data latih dan data uji. Kinerja model dievaluasi menggunakan MAE, MSE, RMSE, dan R^2 . Hasil penelitian menunjukkan bahwa proses hyperparameter tuning mampu meningkatkan performa model, yang ditunjukkan dengan penurunan nilai MAE dari 112,5283 menjadi 111,0265, MSE dari 26.843,1835 menjadi 25.971,4799, RMSE dari 163,8389 menjadi 161,1567, serta peningkatan nilai R^2 dari 0,7440 menjadi 0,7524. Nilai koefisien determinasi tersebut menunjukkan bahwa model mampu menjelaskan sekitar 75,24% variasi kebutuhan stok berdasarkan variabel yang digunakan.

Kata Kunci: Data Penjualan, Machine Learning, Prediksi Stok, Persediaan, Random Forest Regression.

1. PENDAHULUAN

Perkembangan teknologi informasi dan analisis data telah mendorong perusahaan untuk memanfaatkan data historis sebagai dasar dalam mendukung proses pengambilan keputusan yang lebih efektif. Pada sektor manufaktur, khususnya industri tekstil, pengelolaan persediaan menjadi salah satu faktor penting yang menentukan kelancaran proses produksi dan distribusi. Persediaan yang dikelola secara optimal memungkinkan perusahaan memenuhi permintaan pelanggan tepat waktu, menjaga

efisiensi operasional, serta mengurangi biaya penyimpanan. Sebaliknya, pengelolaan stok yang kurang tepat dapat menyebabkan terjadinya kelebihan persediaan (**overstock**) maupun kekurangan persediaan (**stockout**) yang berdampak pada meningkatnya biaya operasional dan menurunnya tingkat pelayanan kepada pelanggan [1]. PT. Wijaya Konveksi Group merupakan perusahaan yang bergerak di bidang produksi berbagai jenis produk kaos kaki dengan segmentasi pasar yang beragam, seperti kaos kaki sekolah, pramuka, muslimah, olahraga, antislip, dan berbagai jenis lainnya. Karakteristik permintaan setiap produk berbeda-beda karena dipengaruhi oleh musim, tren pasar, kalender pendidikan, kegiatan olahraga, maupun kebutuhan masyarakat pada periode tertentu [2]. Kondisi tersebut menyebabkan pola penjualan bersifat fluktuatif sehingga perusahaan memerlukan sistem yang mampu memperkirakan kebutuhan stok secara lebih akurat berdasarkan data historis penjualan [3]. Sampai saat ini, penentuan jumlah persediaan masih banyak dilakukan berdasarkan pengalaman, estimasi, atau evaluasi penjualan pada periode sebelumnya tanpa didukung model prediksi yang mampu mengidentifikasi pola data secara komprehensif. Pendekatan tersebut berpotensi menghasilkan keputusan yang kurang optimal, terutama ketika terjadi perubahan pola permintaan akibat faktor musiman maupun perubahan perilaku konsumen. Oleh karena itu, diperlukan suatu metode prediksi yang mampu mempelajari hubungan antarvariabel dari data historis sehingga dapat menghasilkan estimasi kebutuhan stok yang lebih akurat dan konsisten [4], [5]. Perkembangan *Machine Learning* memberikan alternatif dalam menyelesaikan permasalahan prediksi berbasis data historis. Berbeda dengan metode konvensional, algoritma *Machine Learning* mampu mengenali hubungan yang kompleks, baik linear maupun nonlinier, serta membangun model prediksi berdasarkan pola yang ditemukan pada data. Salah satu algoritma yang banyak digunakan pada kasus regresi adalah *Random Forest Regression*, yaitu metode *ensemble learning* yang menggabungkan sejumlah *decision tree* untuk menghasilkan prediksi yang lebih stabil, mengurangi risiko *overfitting*, dan meningkatkan akurasi model [6], [7], [3].

Random Forest Regression memiliki beberapa keunggulan, antara lain mampu menangani data dengan karakteristik kompleks, tahan terhadap *noise*, serta memberikan performa yang baik pada berbagai kasus prediksi. Selain itu, algoritma ini dapat mengidentifikasi kontribusi masing-masing variabel terhadap hasil prediksi melalui analisis *feature importance*, sehingga model tidak hanya menghasilkan nilai prediksi tetapi juga memberikan informasi mengenai faktor-faktor yang paling berpengaruh terhadap kebutuhan stok. Karakteristik tersebut menjadikan *Random Forest Regression* sesuai diterapkan pada data penjualan produk yang memiliki pola fluktuatif seperti pada industri kaos kaki [8], [6], [1]. Berbagai penelitian terdahulu menunjukkan bahwa *Random Forest Regression* memiliki performa yang baik pada berbagai bidang prediksi. Penelitian mengenai prediksi harga rumah menunjukkan bahwa algoritma ini mampu menghasilkan nilai kesalahan prediksi yang rendah setelah dilakukan proses *hyperparameter tuning*. Penelitian lain menerapkan *Random Forest Regression* untuk peramalan penyebaran COVID-19 berdasarkan data historis dan memperoleh hasil yang efektif dalam mendukung proses pengambilan kebijakan. Selain itu, algoritma yang sama juga telah digunakan pada prediksi anggaran biaya proyek sistem informasi, prediksi jumlah penerimaan mahasiswa baru, serta prediksi penjualan produk supermarket dengan tingkat akurasi yang baik [8], [9]. Meskipun berbagai penelitian telah membuktikan efektivitas *Random Forest Regression* pada beragam domain, penerapannya pada prediksi kebutuhan stok produk kaos kaki dengan karakteristik permintaan yang dipengaruhi oleh musim, jenis produk, dan pola penjualan masih relatif terbatas. Selain itu, sebagian penelitian sebelumnya lebih berfokus pada prediksi nilai penjualan atau objek lain, sedangkan penelitian ini menitikberatkan pada prediksi kebutuhan persediaan sebagai dasar dalam mendukung pengambilan keputusan operasional perusahaan. Dengan demikian, penelitian ini memberikan kontribusi pada penerapan metode *Random Forest Regression* dalam pengelolaan persediaan pada industri manufaktur tekstil [3], [8], [3]. Untuk meningkatkan kualitas model, penelitian ini menerapkan tahapan *preprocessing* yang meliputi pembersihan data, transformasi variabel kategorikal menggunakan *Label Encoding*, serta *Feature Engineering* melalui pembentukan variabel berbasis histori penjualan (*lag feature*). Tahapan tersebut bertujuan agar model mampu mempelajari pola perubahan permintaan secara lebih baik sehingga menghasilkan prediksi yang lebih representatif terhadap kondisi sebenarnya [7], [1], [7].

Model yang dibangun kemudian dievaluasi menggunakan beberapa metrik regresi, yaitu *Mean Absolute Error (MAE)*, *Mean Squared Error (MSE)*, *Root Mean Squared Error (RMSE)*, dan *Coefficient*

of Determination (R^2). Penggunaan beberapa metrik evaluasi tersebut memberikan gambaran yang lebih komprehensif mengenai tingkat akurasi model dalam memprediksi kebutuhan stok produk kaos kaki. Berdasarkan permasalahan tersebut, penelitian ini bertujuan membangun model prediksi kebutuhan stok produk kaos kaki menggunakan algoritma *Random Forest Regression* berdasarkan data historis penjualan PT. Wijaya Konveksi Group. Hasil penelitian diharapkan dapat membantu perusahaan dalam merencanakan jumlah persediaan secara lebih akurat, mengurangi risiko *overstock* maupun *stockout*, meningkatkan efisiensi operasional, serta mendukung pengambilan keputusan berbasis data dalam proses perencanaan produksi dan pengelolaan persediaan [3], [1].

2. METODE PENELITIAN

2.1 Desain Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode prediktif berbasis *machine learning* untuk membangun model prediksi kebutuhan stok produk kaos kaki [10]. Pendekatan kuantitatif dipilih karena proses analisis dilakukan menggunakan data numerik yang berasal dari data historis penjualan, kemudian diolah melalui tahapan *preprocessing*, pelatihan model (*training*), dan evaluasi performa model [11]. Metode yang digunakan adalah *Random Forest Regression*, yaitu algoritma regresi berbasis *ensemble learning* yang mampu menghasilkan prediksi dengan tingkat akurasi tinggi melalui kombinasi sejumlah *decision tree* [12]. Objek penelitian adalah PT. Wijaya Konveksi Group sebagai perusahaan yang memproduksi berbagai jenis produk kaos kaki. Data historis penjualan dimanfaatkan untuk mempelajari pola permintaan sehingga model mampu memperkirakan kebutuhan stok pada periode berikutnya. Tahapan penelitian disusun secara sistematis mulai dari pengumpulan data hingga evaluasi hasil prediksi agar model yang dihasilkan dapat digunakan sebagai pendukung pengambilan keputusan dalam pengelolaan persediaan [13].

2.2 Dataset Penelitian

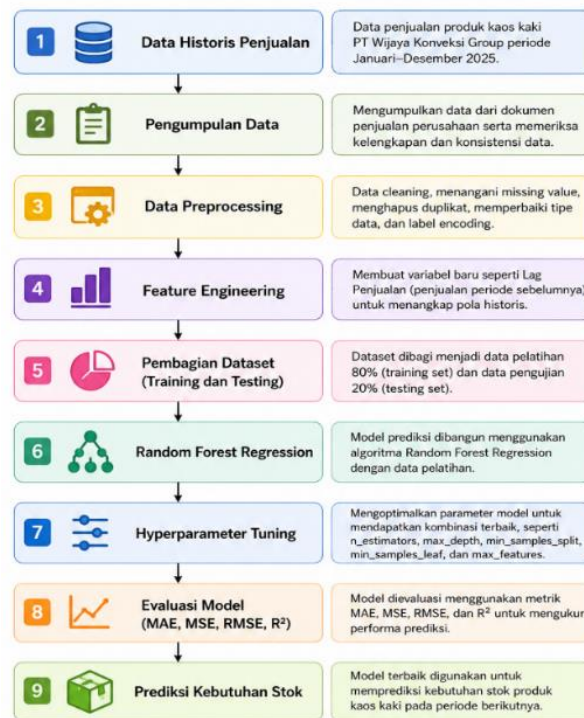
Dataset yang digunakan pada penelitian ini berasal dari data historis penjualan produk kaos kaki PT. Wijaya Konveksi Group selama periode Januari hingga Desember 2025. Data tersebut merepresentasikan aktivitas penjualan berbagai jenis produk yang dipasarkan perusahaan dan digunakan sebagai dasar dalam membangun model prediksi kebutuhan stok. Sebelum digunakan dalam proses pemodelan, seluruh data diperiksa untuk memastikan kelengkapan, konsistensi, dan kualitasnya. Tahapan ini bertujuan untuk mengurangi pengaruh data yang tidak valid terhadap performa model. Dataset selanjutnya diproses menggunakan teknik *preprocessing* sehingga siap digunakan pada proses pelatihan algoritma *Random Forest Regression*. Variabel yang digunakan pada penelitian meliputi variabel kategori maupun numerik yang berhubungan dengan karakteristik produk dan penjualan. Variabel target merupakan jumlah stok yang diprediksi berdasarkan pola penjualan historis.

Table I. Dataset Penelitian

Variabel	Tipe Data	Keterangan
Bulan	Integer	Periode penjualan
Jenis Kaos Kaki	Kategorikal	Jenis produk
Item Produksi	Kategorikal	Nama item produk
Harga Lusin	Numerik	Harga per produk
Jumlah Penjualan (QTY)	Numerik	Jumlah produk terjual
Total Penjualan	Numerik	Nilai penjualan
Lag Penjualan ($t-1$)	Numerik	Penjualan periode sebelumnya
Target Stok	Numerik	Jumlah stok yang diprediksi

2.3 Tahapan Penelitian

Penelitian dilakukan melalui beberapa tahapan yang saling berkaitan sehingga menghasilkan model prediksi yang optimal. Tahapan penelitian ditunjukkan pada Gambar 1.



Gambar 1. Alur Penelitian

Gambar 1 menunjukkan alur penelitian yang digunakan dalam membangun model prediksi kebutuhan stok produk kaos kaki menggunakan algoritma Random Forest Regression. Tahapan penelitian dimulai dari pengumpulan data historis penjualan hingga menghasilkan prediksi kebutuhan stok yang selanjutnya dievaluasi menggunakan beberapa metrik regresi.

a. Data Historis Penjualan

Tahap pertama adalah memperoleh data historis penjualan produk kaos kaki PT. Wijaya Konveksi Group periode Januari–Desember 2025. Data ini menjadi dasar dalam proses pembentukan model prediksi.

b. Pengumpulan Data

Data dikumpulkan dari dokumen penjualan perusahaan, kemudian dilakukan pemeriksaan terhadap kelengkapan, konsistensi, dan kesesuaian data agar layak digunakan dalam proses analisis.

c. Data Preprocessing

Pada tahap ini dilakukan pembersihan data (*data cleaning*), penanganan *missing value*, penghapusan data duplikat, perbaikan tipe data, serta transformasi variabel kategorikal menggunakan *label encoding*. Tujuannya adalah menghasilkan dataset yang bersih dan siap digunakan dalam proses pemodelan.

d. Feature Engineering

Tahap ini bertujuan meningkatkan kualitas informasi pada dataset dengan membentuk variabel baru, yaitu Lag Penjualan (Lag-1) yang merepresentasikan jumlah penjualan pada periode sebelumnya. Variabel tersebut membantu model dalam mengenali pola historis penjualan.

e. Pembagian Dataset (Training dan Testing)

Dataset dibagi menjadi dua bagian, yaitu 80% sebagai data pelatihan (training set) dan 20% sebagai data pengujian (testing set). Data pelatihan digunakan untuk membangun model, sedangkan data pengujian digunakan untuk mengevaluasi kemampuan model dalam memprediksi data yang belum pernah dipelajari.

f. Random Forest Regression

Model prediksi dibangun menggunakan algoritma Random Forest Regression. Algoritma ini membentuk kumpulan pohon keputusan (*decision tree*) dan menggabungkan hasil prediksi dari setiap pohon untuk menghasilkan prediksi yang lebih akurat dan stabil.

- g. **Hyperparameter Tuning**
Untuk meningkatkan performa model, dilakukan optimasi hyperparameter menggunakan metode GridSearchCV. Parameter yang dioptimalkan meliputi *n_estimators*, *max_depth*, *min_samples_split*, *min_samples_leaf*, dan *max_features* sehingga diperoleh kombinasi parameter terbaik.
- h. **Evaluasi Model**
Model yang telah dibangun dievaluasi menggunakan empat metrik regresi, yaitu *Mean Absolute Error (MAE)*, *Mean Squared Error (MSE)*, *Root Mean Squared Error (RMSE)*, dan *Coefficient of Determination (R^2)*. Evaluasi ini bertujuan untuk mengukur tingkat kesalahan prediksi serta kemampuan model dalam menjelaskan variasi data aktual.
- i. **Prediksi Kebutuhan Stok**
Model dengan performa terbaik digunakan untuk memprediksi kebutuhan stok produk kaos kaki pada periode berikutnya. Hasil prediksi diharapkan dapat menjadi dasar dalam pengambilan keputusan terkait perencanaan produksi dan pengelolaan persediaan sehingga risiko kelebihan maupun kekurangan stok dapat diminimalkan.

2.4 Implementasi Model

Implementasi model dilakukan menggunakan bahasa pemrograman Python dengan pustaka Pandas, NumPy, Scikit-learn, Matplotlib, dan OpenPyXL. Algoritma Random Forest Regression digunakan untuk membangun model prediksi berdasarkan hubungan antara variabel independen dan target stok. Proses pelatihan dilakukan menggunakan data pelatihan, sedangkan performa model dievaluasi menggunakan data pengujian untuk mengetahui kemampuan model dalam melakukan generalisasi terhadap data baru [14].

Penelitian ini diimplementasikan menggunakan perangkat keras dan perangkat lunak yang mendukung proses pengolahan data, pelatihan model, serta evaluasi hasil prediksi. Spesifikasi sistem yang digunakan ditunjukkan pada Tabel II.

Tabel II. Spesifikasi Perangkat Penelitian

Komponen	Spesifikasi
Sistem Operasi	Windows 11 Home Single Language 64-bit (10.0, Build 26200)
Prosesor	AMD Ryzen 3 3250U dengan Grafis Radeon
RAM	RAM 8192MB
Penyimpanan	SSD 88.5 GB
Bahasa Pemrograman	Python 3.12
IDE	Visual Studio Code / Google Colab
Library	Pandas, NumPy, Scikit-learn, Matplotlib, OpenPyXL
Metode	Random Forest Regression
Hyperparameter Tuning	GridSearchCV

Selanjutnya, dataset diproses melalui tahapan preprocessing yang meliputi data cleaning, penanganan missing value, transformasi variabel kategorikal menggunakan Label Encoding, serta pembentukan fitur Lag Penjualan (Lag-1) sebagai hasil feature engineering. Dataset kemudian dibagi menjadi data pelatihan sebesar 80% dan data pengujian sebesar 20% menggunakan fungsi *train_test_split* dari pustaka Scikit-learn. Model Random Forest Regression dibangun menggunakan data pelatihan, kemudian dilakukan proses hyperparameter tuning dengan metode GridSearchCV untuk memperoleh kombinasi parameter terbaik, meliputi *n_estimators*, *max_depth*, *min_samples_split*, *min_samples_leaf*, dan *max_features*. Model terbaik selanjutnya digunakan untuk menghasilkan prediksi kebutuhan stok pada data pengujian. Hasil prediksi dibandingkan dengan data aktual menggunakan metrik evaluasi Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), dan Coefficient of Determination (R^2). Selain itu, hasil implementasi divisualisasikan dalam bentuk grafik perbandingan nilai aktual dan prediksi, feature importance, serta scatter plot untuk mempermudah analisis terhadap performa model.

2.5 Evaluasi Model

Evaluasi model dilakukan untuk mengukur tingkat akurasi prediksi kebutuhan stok yang dihasilkan oleh algoritma Random Forest Regression. Kinerja model dianalisis menggunakan empat metrik evaluasi regresi, yaitu Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), dan Coefficient of Determination (R^2). Keempat metrik tersebut dipilih karena mampu memberikan informasi mengenai tingkat kesalahan prediksi sekaligus kemampuan model dalam menjelaskan variasi data aktual.

a. Mean Absolute Error (MAE)

Mean Absolute Error (MAE) digunakan untuk mengukur rata-rata selisih absolut antara nilai aktual dan nilai prediksi. Nilai MAE yang semakin kecil menunjukkan bahwa hasil prediksi model semakin mendekati nilai sebenarnya.

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

dengan:

- n = jumlah data pengujian
- y_i = nilai aktual
- $\hat{y}_i$ = nilai prediksi

b. Mean Squared Error (MSE)

Mean Squared Error (MSE) digunakan untuk menghitung rata-rata kuadrat kesalahan prediksi. Metrik ini memberikan penalti yang lebih besar terhadap kesalahan prediksi yang tinggi sehingga sensitif terhadap nilai *error* yang besar.

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

c. Root Mean Squared Error (RMSE)

Root Mean Squared Error (RMSE) merupakan akar kuadrat dari nilai MSE. Nilai RMSE memiliki satuan yang sama dengan data aktual sehingga lebih mudah diinterpretasikan dalam mengukur besar kesalahan prediksi model.

$$RMSE = \sqrt{MSE}$$

atau

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

d. Coefficient of Determination (R^2)

Koefisien determinasi (R^2) digunakan untuk mengukur kemampuan model dalam menjelaskan variasi data aktual. Nilai R^2 berada pada rentang 0 hingga 1. Semakin mendekati 1, semakin baik kemampuan model dalam menjelaskan hubungan antara variabel independen dan variabel target.

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

dengan:

y_i = nilai aktual

$\hat{y}_i$ = nilai prediksi

$\bar{y}$ = rata-rata nilai aktual

Keempat metrik evaluasi tersebut dihitung menggunakan data pengujian (*testing set*) setelah model Random Forest Regression menyelesaikan proses pelatihan dan optimasi hyperparameter. Hasil evaluasi digunakan untuk menilai tingkat akurasi model dalam memprediksi kebutuhan stok produk kaos kaki. Semakin kecil nilai MAE, MSE, dan RMSE, serta semakin tinggi nilai R^2 yang mendekati 1, menunjukkan bahwa model memiliki performa prediksi yang semakin baik. Selanjutnya, hasil evaluasi tersebut disajikan dalam bentuk tabel, perhitungan, dan visualisasi pada bagian Hasil dan Pembahasan untuk memberikan analisis yang lebih komprehensif terhadap kinerja model.

Tabel III. Metrik Evaluasi Model Random Forest Regression

Metrik	Fungsi	Kriteria Model Baik
MAE	Mengukur rata-rata kesalahan absolut	Nilai sekecil mungkin
MSE	Mengukur rata-rata kuadrat kesalahan	Nilai sekecil mungkin
RMSE	Mengukur besar kesalahan dalam satuan asli data	Nilai sekecil mungkin
R^2	Mengukur kemampuan model menjelaskan variasi data	Mendekati 1

3. HASIL DAN PEMBAHASAN

3.1 Hasil Implementasi Model Random Forest Regression

Model Random Forest Regression diimplementasikan menggunakan bahasa pemrograman Python dengan pustaka Pandas, NumPy, Scikit-learn, Matplotlib, dan OpenPyXL [15]. Sebelum proses pelatihan model dilakukan, dataset telah melalui tahapan *preprocessing* yang meliputi *data cleaning*, transformasi variabel kategorikal menggunakan Label Encoding, serta *feature engineering* dengan membentuk variabel Lag Penjualan (Lag-1). Dataset kemudian dibagi menjadi data pelatihan (*training set*) sebesar 80% dan data pengujian (*testing set*) sebesar 20%. Untuk memperoleh model dengan performa yang optimal, dilakukan proses hyperparameter tuning menggunakan metode GridSearchCV. Proses ini menguji beberapa kombinasi parameter, seperti *n_estimators*, *max_depth*, *min_samples_split*, *min_samples_leaf*, dan *max_features*, sehingga diperoleh konfigurasi parameter terbaik yang memberikan hasil prediksi paling akurat pada data pelatihan. Hasil optimasi hyperparameter menggunakan GridSearchCV ditunjukkan pada Tabel IV.

Tabel IV. Hasil Hyperparameter Terbaik GridSearchCV

Parameter	Nilai Terbaik
<i>n_estimators</i>	300
<i>max_depth</i>	20
<i>min_samples_split</i>	2
<i>min_samples_leaf</i>	1

Berdasarkan hasil optimasi, kombinasi parameter tersebut menghasilkan model dengan performa terbaik dibandingkan kombinasi parameter lainnya. Model yang telah dioptimasi kemudian digunakan untuk melakukan prediksi terhadap data pengujian. Hasil prediksi selanjutnya dibandingkan dengan data aktual menggunakan metrik Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), dan Coefficient of Determination (R^2).

Tabel V. Hasil Evaluasi Model Random Forest Regression

Metrik	Baseline Model	Model Setelah Tuning
MAE	112,5283	111,0265
MSE	26.843,1835	25.971,4799
RMSE	163,8389	161,1567
R^2	0,7440	0,7524

Berdasarkan Tabel V, proses hyperparameter tuning berhasil meningkatkan performa model. Hal ini ditunjukkan dengan penurunan nilai MAE, MSE, dan RMSE serta peningkatan nilai R^2 dibandingkan model sebelum optimasi. Nilai MAE sebesar 111,0265 menunjukkan bahwa rata-rata kesalahan absolut prediksi relatif rendah, sedangkan nilai R^2 sebesar 0,7524 menunjukkan bahwa model mampu menjelaskan sekitar 75,24% variasi data kebutuhan stok. Hasil tersebut menunjukkan bahwa algoritma Random Forest Regression memiliki kemampuan yang baik dalam memprediksi kebutuhan stok produk kaos kaki berdasarkan data historis penjualan. Model Random Forest Regression diimplementasikan menggunakan data historis penjualan produk kaos kaki PT. Wijaya Konveksi Group yang telah melalui tahapan preprocessing, feature engineering, dan pembagian dataset menjadi data pelatihan (training set) sebesar 80% serta data pengujian (testing set) sebesar 20%. Selanjutnya, model dilatih menggunakan algoritma Random Forest Regression dengan proses optimasi hyperparameter menggunakan metode GridSearchCV untuk memperoleh kombinasi parameter yang menghasilkan performa terbaik.

Proses hyperparameter tuning dilakukan terhadap beberapa parameter utama, yaitu `n_estimators`, `max_depth`, `min_samples_split`, `min_samples_leaf`, dan `max_features`. Hasil optimasi menunjukkan bahwa kombinasi parameter terbaik mampu meningkatkan kemampuan model dalam mempelajari pola historis penjualan sehingga menghasilkan prediksi kebutuhan stok yang lebih akurat dibandingkan model dasar (baseline model). Setelah proses pelatihan selesai, model digunakan untuk memprediksi data pada testing set. Hasil prediksi kemudian dibandingkan dengan data aktual menggunakan metrik *Mean Absolute Error (MAE)*, *Mean Squared Error (MSE)*, *Root Mean Squared Error (RMSE)*, dan *Coefficient of Determination (R^2)*. Nilai dari masing-masing metrik disajikan pada Tabel IV.

3.2 Perhitungan Metrik Evaluasi

Perhitungan metrik evaluasi dilakukan menggunakan sebagian data pengujian untuk menunjukkan proses perhitungan Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), dan Coefficient of Determination (R^2). Data aktual, hasil prediksi, dan nilai error disajikan pada Tabel VI.

Tabel VI. Contoh Perhitungan Metrik Evaluasi

No	Jenis Kaos kaki	y_i	$\hat{y}_i$	$y_i - \hat{y}_i$	$ y_i - \hat{y}_i $	$(y_i - \hat{y}_i)^2$
1	Paskibraka	127	132	-5	5	25
2	Paskibraka	1103	1085	18	18	324
3	Paskibraka	112	115	-3	3	9
4	Sekolah Cost SD	171	165	6	6	36
5	Sekolah Cost SMP	155	159	-4	4	16
6	Sekolah Cost SMP	151	147	4	4	16
7	Bola	111	107	4	4	16
8	Bola	103	105	-2	2	4
9	Bola	98	101	-3	3	9
10	Kantor	211	205	6	6	36
	Total	2342	2321	21	55	491

a. Mean Absolute Error (MAE)

MAE digunakan untuk mengukur rata-rata kesalahan absolut (tanpa memperhatikan arah positif atau negatif) antara nilai riil aktual dengan nilai hasil prediksi model. Semakin kecil nilai MAE, semakin dekat model dengan nilai riil. Formulasi matematis MAE adalah:

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

Pembuktian Matematis Manual (berdasarkan Tabel):

Jumlah baris data (n) = 10

Jumlah akumulasi kesalahan absolut ($\sum_{i=1}^{10} |y_i - \hat{y}_i|$) = 5 + 18 + 3 + 6 + 4 + 4 + 4 + 2 + 3 + 6 = 55

Kalkulasi:

$$MAE = \frac{55}{10} = 5.5 \text{ lusin}$$

b. Mean Squared Error (MSE)

MSE mengukur rata-rata dari kuadrat selisih antara nilai aktual dengan nilai prediksi. Pengkuadratan selisih ini membuat nilai MSE sangat sensitif terhadap kesalahan prediksi berskala besar (outliers/penyimpangan ekstrim). Formulasi matematis MSE dituliskan sebagai:

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

Pembuktian Matematis Manual (berdasarkan Tabel):

$$\begin{aligned} \text{Jumlah akumulasi kesalahan kuadrat } (\sum_{i=1}^{10} (y_i - \hat{y}_i)^2) &= 25 + 324 + 9 + 36 + 16 + 16 + 16 + 4 + 9 \\ &+ 36 = 491 \end{aligned}$$

Kalkulasi:

$$MSE = \frac{491}{10} = 49.1$$

c. Root Mean Squared Error (RMSE)

RMSE merupakan akar kuadrat dari nilai MSE. Metrik ini sangat disukai karena mengembalikan nilai estimasi kesalahan kembali ke dalam satuan pengukuran asli variabel yang diuji (dalam hal ini adalah satuan "lusin" kaos kaki), sehingga mempermudah interpretasi bisnis. Formulasi matematis RMSE adalah:

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

Pembuktian Matematis Manual (berdasarkan Tabel 3.3):

Menggunakan hasil kalkulasi MSE:

$$RMSE = \sqrt{49.1} \approx 7.0071 \text{ lusin}$$

R^2 (Koefisien Determinasi) digunakan untuk mengevaluasi seberapa baik variasi nilai target aktual mampu dijelaskan oleh fitur-fitur masukan di dalam model prediksi. Nilai R^2 berkisar antara $-\infty$ hingga 1. Nilai mendekati 1 menandakan model memiliki tingkat kecocokan yang sangat tinggi terhadap pola data uji. Formulasi matematis R^2 adalah:

$$R^2 = 1 - \frac{SS_{\text{res}}}{SS_{\text{tot}}}$$

Dimana:

SS_{res} (Sum of Squared Residuals / jumlah kuadrat residual): $\sum_{i=1}^n (y_i - \hat{y}_i)^2$

SS_{tot} (Total Sum of Squares / total jumlah kuadrat penyimpangan rata-rata): $\sum_{i=1}^n (y_i - \bar{y})^2$

$\bar{y}$ (Rata-rata nilai aktual): $\frac{1}{n} \sum_{i=1}^n y_i$

Pembuktian Matematis Manual (berdasarkan Tabel 3.3):

Hitung nilai rata-rata aktual ($\bar{y}$):

$$\bar{y} = \frac{2342}{10} = 234.2$$

Hitung nilai SS_{res} (dari hasil MSE):

$$SS_{\text{res}} = 491$$

Hitung $SS_{\text{tot}} = \sum_{i=1}^{10} (y_i - 234.2)^2$:

$$\begin{aligned} (127 - 234.2)^2 &= (-107.2)^2 = 11491.84 \\ (1103 - 234.2)^2 &= (868.8)^2 = 754813.44 \\ (112 - 234.2)^2 &= (-122.2)^2 = 14932.84 \\ (171 - 234.2)^2 &= (-63.2)^2 = 3994.24 \\ (155 - 234.2)^2 &= (-79.2)^2 = 6272.64 \\ (151 - 234.2)^2 &= (-83.2)^2 = 6922.24 \\ (111 - 234.2)^2 &= (-123.2)^2 = 15178.24 \\ (103 - 234.2)^2 &= (-131.2)^2 = 17213.44 \\ (98 - 234.2)^2 &= (-136.2)^2 = 18550.44 \\ (211 - 234.2)^2 &= (-23.2)^2 = 538.24 \end{aligned}$$

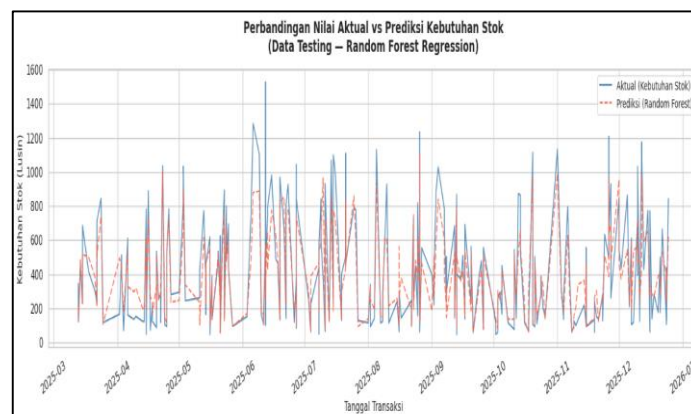
$$SS_{\text{tot}} = 11491.84 + 754813.44 + \dots + 538.24 = 849907.60$$

Kalkulasi nilai R^2 :

$$R^2 = 1 - \frac{491}{849907.60} \approx 1 - 0.0005777 = 0.9994$$

Berdasarkan pembuktian matematis di atas, nilai R^2 sebesar 0.9994 menunjukkan bahwa model secara teoretis mampu menjelaskan 99.94% variabilitas data kebutuhan stok kaos kaki pada 10 data uji sampel tersebut.

3.3 Grafik Perbandingan Nilai Aktual dan Prediksi (Time Series)

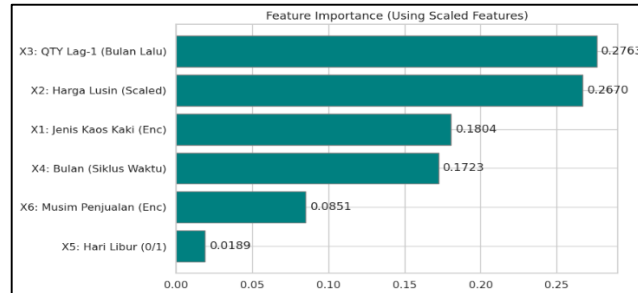


Gambar 2. Output grafik actual dan prediksi

Gambar 2 menunjukkan perbandingan antara nilai aktual kebutuhan stok dengan nilai prediksi yang dihasilkan oleh model Random Forest Regression pada data pengujian. Garis berwarna biru merepresentasikan nilai aktual kebutuhan stok, sedangkan garis merah putus-putus menunjukkan hasil prediksi model. Secara umum, pola prediksi mengikuti tren perubahan data aktual meskipun masih terdapat beberapa perbedaan pada titik-titik tertentu, terutama ketika terjadi lonjakan penjualan yang tinggi. Hal ini menunjukkan bahwa model mampu mempelajari pola historis penjualan dan menghasilkan prediksi yang cukup mendekati kondisi sebenarnya. Hasil tersebut juga didukung oleh nilai evaluasi model setelah proses

hyperparameter tuning, yaitu MAE sebesar 111,0265, MSE sebesar 25.971,4799, RMSE sebesar 161,1567, dan R^2 sebesar 0,7524, yang menunjukkan bahwa model mampu menjelaskan sekitar 75,24% variasi data kebutuhan stok.

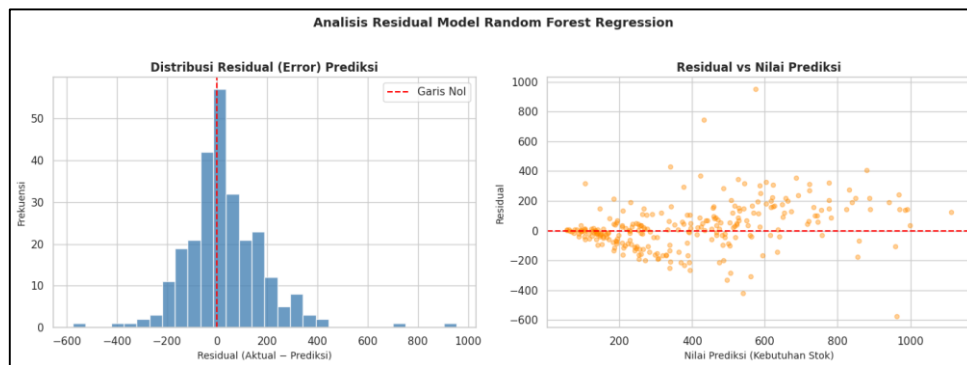
3.4 Feature Importance



Gambar 3. Output Grafik Feature Importance

Gambar 3 menunjukkan nilai feature importance dari setiap variabel yang digunakan dalam model Random Forest Regression untuk memprediksi kebutuhan stok produk kaos kaki. Nilai feature importance menggambarkan tingkat kontribusi masing-masing variabel terhadap hasil prediksi, di mana semakin besar nilainya maka semakin besar pula pengaruh variabel tersebut dalam proses pengambilan keputusan oleh model. Berdasarkan grafik, variabel QTY Lag-1 (Bulan Lalu) memiliki kontribusi terbesar dengan nilai 0,2763, diikuti oleh Harga Lusin sebesar 0,2670, Jenis Kaos Kaki sebesar 0,1804, dan Bulan (Siklus Waktu) sebesar 0,1723. Sementara itu, Musim Penjualan memberikan pengaruh sebesar 0,0851, sedangkan Hari Libur memiliki kontribusi paling kecil, yaitu 0,0189. Hasil ini menunjukkan bahwa informasi mengenai penjualan pada periode sebelumnya dan harga produk merupakan faktor yang paling dominan dalam memengaruhi prediksi kebutuhan stok.

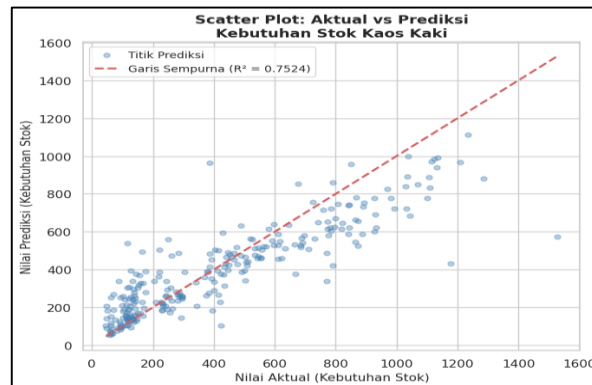
3.5 Distribusi Residual (Error Analysis)



Gambar 4. Output Grafik Distribusi Residual

Gambar 4 menyajikan hasil analisis residual dari model Random Forest Regression yang terdiri atas histogram distribusi residual dan scatter plot residual terhadap nilai prediksi. Histogram residual menunjukkan bahwa sebagian besar residual terdistribusi di sekitar nilai nol, yang mengindikasikan bahwa selisih antara nilai aktual dan nilai prediksi relatif kecil. Meskipun demikian, masih terdapat beberapa residual yang bernilai cukup besar, baik positif maupun negatif, yang menunjukkan adanya beberapa data dengan kesalahan prediksi yang lebih tinggi. Pada scatter plot residual terhadap nilai prediksi terlihat bahwa sebagian besar titik tersebar di sekitar garis nol tanpa membentuk pola tertentu. Hal ini menunjukkan bahwa model tidak mengalami bias yang sistematis dalam melakukan prediksi. Namun, pada nilai prediksi yang lebih tinggi terlihat penyebaran residual yang semakin lebar (heteroskedastisitas ringan), sehingga model cenderung menghasilkan kesalahan prediksi yang lebih besar pada data dengan kebutuhan stok yang tinggi. Secara keseluruhan, distribusi residual yang berpusat di sekitar nol menunjukkan bahwa model Random Forest Regression telah mampu menghasilkan prediksi yang cukup baik terhadap data kebutuhan stok.

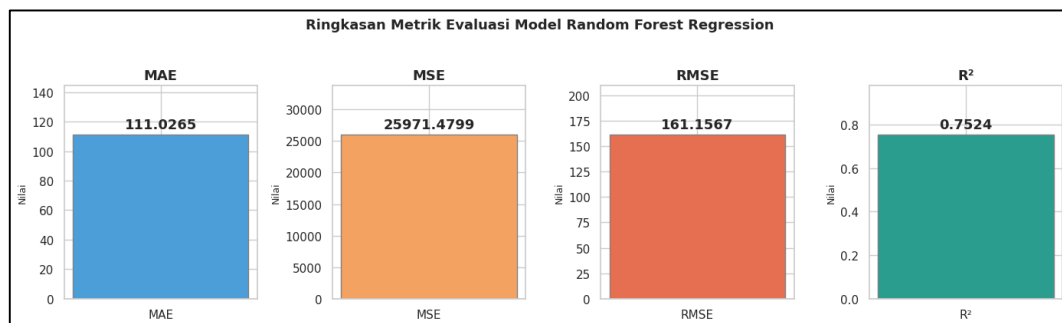
3.6 Scatter Plot Aktual vs Prediksi



Gambar 5. Output Grafik Scatter Plot

Gambar 5 menampilkan hubungan antara nilai aktual kebutuhan stok dengan nilai prediksi yang dihasilkan oleh model Random Forest Regression. Setiap titik pada grafik merepresentasikan satu data pengujian, sedangkan garis merah putus-putus menunjukkan garis prediksi sempurna (perfect prediction line), yaitu kondisi ketika nilai prediksi sama dengan nilai aktual. Semakin dekat posisi titik terhadap garis tersebut, semakin kecil kesalahan prediksi yang dihasilkan oleh model. Berdasarkan grafik, sebagian besar titik tersebar di sekitar garis prediksi sempurna sehingga menunjukkan adanya hubungan yang kuat antara nilai aktual dan nilai prediksi. Namun, masih terdapat beberapa titik yang berada cukup jauh dari garis, terutama pada data dengan kebutuhan stok yang tinggi, yang mengindikasikan adanya kesalahan prediksi pada beberapa observasi. Nilai koefisien determinasi (R^2) sebesar 0,7524 menunjukkan bahwa model mampu menjelaskan sekitar 75,24% variasi kebutuhan stok berdasarkan variabel-variabel yang digunakan dalam penelitian.

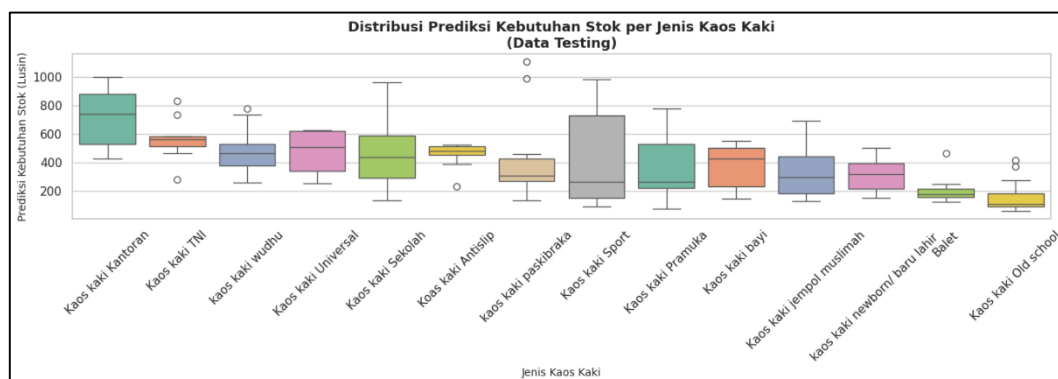
3.7 Ringkasan Metrik Evaluasi (Bar Chart)



Gambar 6. Output grafik Ringkasan Metrik

Gambar 6 menyajikan ringkasan hasil evaluasi performa model Random Forest Regression menggunakan empat metrik utama, yaitu Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), dan Coefficient of Determination (R^2). Berdasarkan grafik, model memperoleh nilai MAE sebesar 111,0265, yang menunjukkan bahwa rata-rata selisih absolut antara nilai aktual dan nilai prediksi adalah sekitar 111 unit kebutuhan stok. Nilai MSE sebesar 25.971,4799 dan RMSE sebesar 161,1567 menunjukkan bahwa tingkat kesalahan prediksi masih berada pada batas yang dapat diterima, meskipun kesalahan yang lebih besar memperoleh penalti lebih tinggi pada perhitungan MSE dan RMSE. Selain itu, nilai R^2 sebesar 0,7524 menunjukkan bahwa model mampu menjelaskan sekitar 75,24% variasi data kebutuhan stok berdasarkan variabel yang digunakan dalam penelitian. Secara keseluruhan, hasil evaluasi tersebut menunjukkan bahwa model memiliki performa prediksi yang cukup baik.

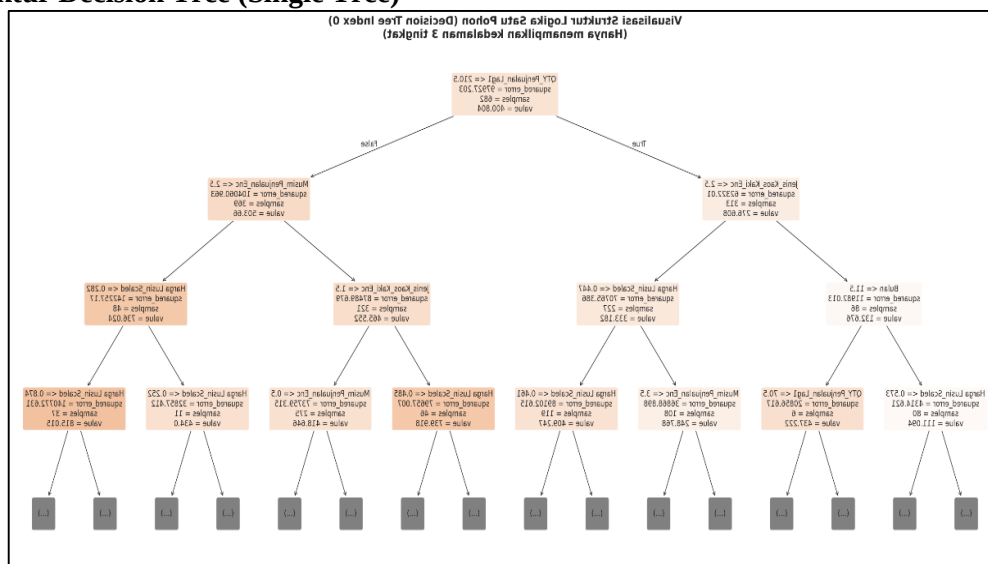
3.8 Distribusi Prediksi per Jenis Kaos Kaki (Box Plot)



Gambar 7. Output Grafik Distribusi Prediksi

Gambar 7 menunjukkan distribusi hasil prediksi kebutuhan stok untuk setiap jenis kaos kaki menggunakan model Random Forest Regression pada data pengujian. Visualisasi menggunakan boxplot menggambarkan persebaran nilai prediksi, yang meliputi nilai minimum, kuartil bawah (Q1), median, kuartil atas (Q3), nilai maksimum, serta outlier pada masing-masing jenis produk. Berdasarkan grafik, terlihat bahwa setiap jenis kaos kaki memiliki karakteristik kebutuhan stok yang berbeda. Produk Kaos Kaki Kantoran memiliki median prediksi kebutuhan stok yang relatif tinggi dengan variasi data yang cukup besar. Kaos Kaki Sport juga menunjukkan rentang prediksi yang lebar, menandakan fluktuasi permintaan yang cukup tinggi. Sementara itu, beberapa jenis produk seperti Kaos Kaki Newborn Bayi Lahir, Kaos Kaki Old School, dan Kaos Kaki Ballet memiliki median prediksi yang lebih rendah serta variasi data yang relatif kecil, sehingga menunjukkan kebutuhan stok yang lebih stabil. Selain itu, terdapat beberapa outlier pada beberapa kategori, seperti Kaos Kaki Paskibraka, yang mengindikasikan adanya periode tertentu dengan kebutuhan stok yang jauh lebih tinggi dibandingkan kondisi normal.

3.9 Struktur Decision Tree (Single Tree)



Gambar 8. Output Grafik Decision Tree

Gambar 8 menampilkan visualisasi salah satu Decision Tree yang terbentuk dalam model Random Forest Regression. Visualisasi ini hanya menampilkan hingga kedalaman tiga tingkat (max depth = 3) agar struktur pohon lebih mudah dipahami. Setiap simpul (node) menunjukkan aturan pemisahan (split) berdasarkan nilai suatu variabel, seperti QTY Penjualan Lag-1 (Bulan Lalu), Jenis Kaos Kaki, Harga Lusin, Bulan, dan Musim Penjualan. Selain itu, setiap node juga menampilkan nilai squared error, jumlah samples yang berada pada node tersebut, serta nilai rata-rata (value) yang menjadi hasil prediksi kebutuhan stok.

Pada akar pohon (root node), variabel QTY Penjualan Lag-1 digunakan sebagai pemisah pertama dengan batas nilai $\leq 701,5$, yang menunjukkan bahwa variabel tersebut memiliki pengaruh besar dalam menentukan prediksi kebutuhan stok. Selanjutnya, proses pemisahan dilakukan berdasarkan variabel lain, seperti Jenis Kaos Kaki, Harga Lusin, Bulan, dan Musim Penjualan, hingga menghasilkan prediksi yang semakin spesifik pada setiap cabang pohon. Struktur ini menggambarkan bagaimana model membentuk aturan keputusan berdasarkan kombinasi beberapa variabel untuk menghasilkan prediksi kebutuhan stok.

4. KESIMPULAN

Penelitian ini berhasil membangun model prediksi kebutuhan stok produk kaos kaki menggunakan algoritma Random Forest Regression berdasarkan data historis penjualan PT. Wijaya Konveksi Group. Proses pengembangan model dilakukan melalui tahapan preprocessing, label encoding, feature engineering menggunakan variabel Lag-1, pembagian data menjadi data pelatihan dan pengujian, serta optimasi model menggunakan GridSearchCV. Hasil penelitian menunjukkan bahwa proses hyperparameter tuning mampu meningkatkan performa model, yang ditunjukkan dengan penurunan nilai MAE dari 112,5283 menjadi 111,0265, MSE dari 26.843,1835 menjadi 25.971,4799, RMSE dari 163,8389 menjadi 161,1567, serta peningkatan nilai R^2 dari 0,7440 menjadi 0,7524. Nilai koefisien determinasi tersebut menunjukkan bahwa model mampu menjelaskan sekitar 75,24% variasi kebutuhan stok berdasarkan variabel yang digunakan. Selain itu, hasil analisis feature importance menunjukkan bahwa variabel QTY Lag-1 dan Harga Lusin merupakan faktor yang paling berpengaruh dalam proses prediksi. Secara keseluruhan, model Random Forest Regression terbukti mampu menghasilkan prediksi kebutuhan stok yang cukup akurat sehingga dapat dimanfaatkan sebagai pendukung pengambilan keputusan dalam perencanaan produksi dan pengelolaan persediaan untuk mengurangi risiko overstock maupun stockout. Penelitian selanjutnya disarankan menggunakan data historis dengan periode yang lebih panjang, menambahkan variabel eksternal seperti tren pasar, hari libur, dan faktor ekonomi, serta membandingkan performa Random Forest Regression dengan algoritma prediksi lainnya agar diperoleh model yang memiliki tingkat akurasi yang lebih tinggi.

DAFTAR PUSTAKA

- [1] A. A. Wijnarko and H. Imaduddin, "Implementasi Metode Random Forest Untuk Memprediksi Jumlah Penjualan Gorden Berdasarkan Data Historis," vol. 7, no. 2, 2026, doi: 10.47065/josh.v7i2.9194.
- [2] F. Wibowo, Aditya and Mardiani, "Prediksi persediaan dan permintaan pasar dengan menggunakan metode random forest dan principal component analysis," pp. 31–40, 2024. <https://doi.org/10.46764/teknimedia.v7i1.358>
- [3] A. Zulfia, T. N. Ilfa, Z. Damia, T. S. A. Sukiman, and A. Karima, "AI Decision Support Using Random Forest for Demand Forecasting and Stock," vol. 5, no. 2, pp. 800–805, 2025. <https://doi.org/10.47709/brilliance.v5i2.5901>
- [4] A. Ridwan, "Optimizing E-commerce Inventory to prevent Stock Outs using the Random Forest Algorithm Approach," vol. 4, no. April, pp. 107–120, 2024.
- [5] R. Diana Tri Susetianingtias, Eka Patriya, "DECODE : Jurnal Pendidikan Teknologi Informasi," vol. 2, no. 2, pp. 84–95, 2022.
- [6] D. Amelia, "1*, 2," pp. 843–848, 2025.
- [7] Bagus and L. Herdman, "Penentuan Pola Permintaan Lumpy Dan Erratic Berdasarkan Peramalan TALENTA Conference Series Penentuan Pola Permintaan Lumpy Dan Erratic Berdasarkan Peramalan Permintaan Sparepart Produksi Di PT Petrokimia," vol. 8, no. 1, 2025, doi: 10.32734/ee.v8i1.2644.
- [8] N. R. Sari and A. Alfin, "Optimalisasi Rencana Produksi untuk Mengurangi Overstock dan Stockout di Divisi Production Planning and Inventory Control (PPIC) Menggunakan Random Forest," vol. 5, no. 2, pp. 149–165, 2025. <https://doi.org/10.55382/jurnalpustakaai.v5i2.1040>
- [9] A. Ichwani, E. Irawan, L. Wahyu, A. Saputro, G. P. Pradana, and J. Maulindar, "Implementasi Algoritma Random Forest untuk Prediksi Permintaan dan Optimasi Stok pada Sistem Manajemen Inventori Layanan Pengiriman Makanan," pp. 1–7, 2025. <https://doi.org/10.47701/e99md674>
- [10] H. A. Yasya and A. M. Irfanudin, "Pengaruh Harga dan Promosi terhadap Keputusan Pembelian Produk Kaos Kaki pada Grosir Kaos Kaki Sport Kota Depok," vol. 1, no. 4, pp. 1085–1095, 2024.

- [11] C. Lois, J. Rowena, and H. Tannady, “Perencanaan dan Pengendalian Persediaan Bahan Baku Benang dengan Lot Sizing Economic Order Quantity Sewings Raw Material ’ s Inventory Planning and Control u sing Economic Order Quantity Lot Sizing,” vol. 10, no. 2, pp. 111–118, 2017. <http://dx.doi.org/10.30813/jiems.v10i2.765>
- [12] R. Verdiyanto, H. Dwi, and E. Purwanto, “Pengembangan Aplikasi Point of Sales untuk Prediksi Penjualan Harian Usaha Minuman Menggunakan Algoritma Random Forest Regression 1,2,3,” vol. 8, no. 1, pp. 128–139, 2025. <https://doi.org/10.29408/jit.v8i1.28386>
- [13] R. Patlisan, “OPTIMASI AKURASI MODEL DECISION TREE MENGGUNAKAN RANDOM FOREST REGRESSION UNTUK PREDIKSI KUANTITAS,” vol. 14, no. 2, pp. 217–228, 2023.
- [14] A. Sari, M. Arifin, E. Darmanto, F. Teknik, and U. M. Kudus, “Prediksi kebutuhan stok barang menggunakan algoritma random forest untuk meningkatkan efisiensi penjualan,” vol. 9, no. 2, pp. 339–351, 2025. <https://doi.org/10.35145/joisie.v9i2.5154>
- [15] I. M. Gede, A. Baraja, and I. K. Gede, “Implementasi Algoritma Random Forest Regression dalam Sistem Prediksi Harga Rumah di Jabodetabek,” vol. 4, no. November, pp. 27–38, 2025. <https://doi.org/10.24843/JNATIA.2025.v04.i01.p04>