



Implementasi Metode Random Forest Untuk Memprediksi Jumlah Penjualan Gorden Berdasarkan Data Historis

Amiladito Adhyatma Wijanarko, Helmi Imaduddin*

Fakultas Komunikasi dan Informatika, Teknik Informatika, Universitas Muhammadiyah Surakarta, Surakarta

Jl. A. Yani, Mendungan, Pabelan, Kec. Kartasura, Kabupaten Sukoharjo, Jawa Tengah, Indonesia

Email: ¹1200220056@student.ums.ac.id, ^{2,*}hi776@ums.ac.id

Email Penulis Korespondensi: hi776@ums.ac.id

Submitted: 12/01/2026; Accepted: 26/01/2026; Published: 28/01/2026

Abstrak—Perkembangan teknologi informasi yang pesat mendorong perusahaan, termasuk usaha kecil Tova Gorden yang bergerak di bidang penjualan gorden untuk mengadopsi teknologi guna meningkatkan efisiensi operasional dan daya saing. Tova Gorden sering menghadapi kendala dalam pemenuhan pesanan, terutama saat permintaan meningkat tiba-tiba, yang diperparah dengan keterbatasan stok, kesulitan bahan baku (seperti smokering), sistem pre-order kain, dan proses produksi yang memakan waktu. Penentuan stok yang masih berdasarkan perkiraan karyawan sering menimbulkan inefisiensi berupa kekurangan atau kelebihan barang. Kondisi ini menyoroti kebutuhan mendesak akan sistem prediksi yang akurat untuk mengoptimalkan manajemen persediaan. Penelitian ini bertujuan untuk mengimplementasikan dan menguji kinerja algoritma Random Forest, yang merupakan metode ensemble learning, untuk memprediksi jumlah penjualan gorden berdasarkan data historis penjualan. Data yang dikumpulkan meliputi informasi historis terkait penjualan gorden yang berisi minggu penjualan, motif gorden, jumlah penjualan. Berbeda dengan penelitian sebelumnya yang umumnya menggunakan Regresi Linier dan berfokus pada produk dengan pola penjualan stabil, penelitian ini menerapkan Random Forest untuk mengatasi pola permintaan gorden yang lebih fluktuatif. Metode penelitian ini mencakup beberapa tahapan, yaitu Pengumpulan data, Exploratory Data Analysis (EDA), Preprocessing Data, Data Splitting (70% latih, 15% validasi, 15% uji), Modeling dengan Random Forest, Evaluasi, dan Deployment. Hasil evaluasi menunjukkan bahwa model memiliki kinerja yang sangat baik, dengan nilai koefisien determinasi (R^2) sebesar 97,83% pada data latih, 93,72% pada data validasi, dan 96,64% pada data uji. Selanjutnya, model tersebut diintegrasikan ke dalam sebuah sistem berbasis web menggunakan framework Flask. Sistem ini dilengkapi dengan fitur unggah data, proses prediksi serta pengelompokan kategori gorden, dan penyajian hasil evaluasi model.

Kata Kunci: Random Forest; Prediksi Penjualan; Manajemen Persediaan; Efisiensi operasional; Algoritma Ensemble

Abstract—The rapid development of information technology has encouraged companies, including Tova Gorden, a small business engaged in curtain sales, to adopt technology to improve operational efficiency and competitiveness. Tova Gorden often faces obstacles in fulfilling orders, especially when demand suddenly increases, which is exacerbated by limited stock, raw material difficulties (such as smokers), fabric pre-order systems, and time-consuming production processes. Determining stock that is still based on employee estimates often leads to inefficiencies in the form of shortages or excesses of goods. This condition highlights the urgent need for an accurate prediction system to optimize inventory management. This study aims to implement and test the performance of the Random Forest algorithm, which is an ensemble learning method, to predict the number of curtain sales based on historical sales data. The collected data includes historical information related to curtain sales, including sales weeks, curtain motifs, and sales volumes. Unlike previous studies that generally use Linear Regression and focus on products with stable sales patterns, this study applies Random Forest to address more fluctuating curtain demand patterns. This research method includes several stages, namely Data Collection, Exploratory Data Analysis (EDA), Data Preprocessing, Data Splitting (70% training, 15% validation, 15% testing), Modeling with Random Forest, Evaluation, and Deployment. The evaluation results show that the model has excellent performance, with a coefficient of determination (R^2) value of 97.83% on training data, 93.72% on validation data, and 96.64% on test data. Furthermore, the model is integrated into a web-based system using the Flask framework. This system is equipped with data upload features, prediction processes and curtain category grouping, and presentation of model evaluation results.

Keywords: Random Forest; Sales Prediction; Inventory Management; Operational Efficiency; Ensemble Algorithm

1. PENDAHULUAN

Pada era digital saat ini, perkembangan teknologi informasi berlangsung dengan sangat cepat dan pesat [1]. Kemajuan ini membuat penyebaran informasi menjadi lebih cepat dan mudah [2]. Perkembangan teknologi informasi yang semakin pesat ternyata memberikan dampak signifikan dalam berbagai bidang, termasuk dunia perdagangan. Perusahaan membutuhkan pemanfaatan teknologi untuk strategi dan analisis bisnis perdagangan [3]. Dengan strategi yang tepat dan analisis data yang akurat, perusahaan dapat mengambil keputusan lebih cepat dan tepat.

Usaha kecil seperti Tova Gorden yang bergerak di bidang penjualan gorden masih menghadapi beberapa kendala dalam kegiatan operasionalnya. Salah satu masalah yang sering muncul adalah ketika permintaan pelanggan tiba-tiba meningkat, Tova Gorden tidak selalu bisa memenuhi pesanan karena stok yang terbatas. Situasi ini makin rumit karena sebagian pelanggan Tova Gorden merupakan penjual di platform TikTok yang memiliki batas waktu pengiriman yang ketat. Selain itu, bahan baku seperti smokering kadang sulit didapat, sementara pembelian kain harus dilakukan dengan sistem pre-order. Proses pembuatan gorden juga memerlukan waktu cukup lama, mulai dari pemotongan kain, penjahitan, hingga pengemasan, sehingga sering terjadi keterlambatan dalam pemenuhan pesanan dan menurunkan efisiensi operasional perusahaan. Penentuan jumlah dan waktu pemesanan stok yang masih berdasarkan perkiraan karyawan sering menimbulkan ketidakefisienan, seperti kekurangan barang

saat permintaan tinggi atau kelebihan barang saat penjualan menurun[4]. Menjaga ketersediaan stok yang tepat waktu dan sesuai dengan permintaan pasar adalah tantangan penting untuk menghindari kekosongan atau kelebihan persediaan yang dapat merugikan [5].

Kondisi tersebut menunjukkan perlunya sistem prediksi yang dapat memperkirakan jumlah penjualan gorden dengan lebih akurat berdasarkan data penjualan sebelumnya. Penggunaan metode Random Forest merupakan salah satu alternatif yang tepat, karena mampu memproses data yang kompleks dan memberikan hasil prediksi yang akurat serta stabil. Penelitian sebelumnya telah memanfaatkan metode Regresi Linier untuk melakukan prediksi penjualan buku serta memperkirakan jumlah persediaan barang pada periode berikutnya [6]. Meskipun Regresi Linier telah terbukti mampu melakukan prediksi penjualan dan persediaan barang, metode ini memiliki keterbatasan dalam menangani data dengan pola yang tidak linear dan kompleks. Selain itu, penelitian-penelitian sebelumnya umumnya masih berfokus pada produk umum seperti buku atau barang konsumsi dengan pola penjualan yang relatif stabil. Hal ini menunjukkan adanya celah penelitian, yaitu belum banyaknya kajian yang menerapkan metode berbasis ensemble learning seperti Random Forest untuk memprediksi penjualan pada bisnis gorden, yang memiliki pola permintaan lebih fluktuatif. Kelebihan dari algoritma Random Forest mencakup tingkat akurasi yang tinggi serta kemampuan untuk mengolah data dengan dimensi yang besar [7].

Berdasarkan permasalahan yang telah diuraikan, penelitian ini memiliki tujuan untuk mengimplementasikan algoritma Random Forest dalam melakukan prediksi jumlah penjualan pada usaha penjualan gorden. Algoritma Random Forest Regressor adalah metode pembelajaran ensemble yang menggabungkan banyak decision tree, dengan cara mengumpulkan prediksi dari setiap pohon untuk menghasilkan hasil akhir yang lebih akurat [8]. Untuk menghasilkan prediksi yang akurat, diperlukan data yang dapat merepresentasikan kondisi penjualan sebelumnya. Oleh karena itu, data historis penjualan yang dimiliki dapat dimanfaatkan melalui metode analisis untuk menemukan pola penjualan tertentu, yang kemudian digunakan sebagai dasar dalam melakukan peramalan atau prediksi penjualan di periode berikutnya [9]. Data historis penjualan yang dikumpulkan meliputi informasi terkait minggu penjualan, motif gorden, serta jumlah penjualan. Tujuan dari pengenalan pola ini adalah untuk memperkirakan jumlah penjualan yang akan terjadi serta mengidentifikasi segmen pasar yang berpotensi mendominasi di masa mendatang. Melalui metode yang diterapkan, diharapkan Tova Gorden dapat mengoptimalkan keputusan pemesanan stok dengan cara yang lebih terencana dan sesuai dengan dinamika permintaan pelanggan.

2. METODOLOGI PENELITIAN

Metode penelitian merupakan pendekatan ilmiah yang digunakan untuk memperoleh data dengan tujuan tertentu [10]. Penelitian ini bertujuan untuk menguji kinerja algoritma Random Forest dalam memprediksi jumlah penjualan gorden berdasarkan data historis penjualan pada Tova Gorden. Algoritma Random Forest dipilih karena memiliki keunggulan dalam mengolah data dengan banyak variabel serta mampu menghasilkan prediksi yang konsisten dan akurat. Penelitian ini terdiri dari beberapa tahapan yang harus dilalui. Tahapan tersebut dapat dilihat pada Gambar 1.



Gambar 1. Tahapan Penelitian

2.1 Pengumpulan Data

Pada tahap ini pengumpulan data dilakukan pada Tova Gorden. Proses pengumpulan data dalam penelitian ini melalui beberapa tahap yang cermat [11]. Data yang dikumpulkan meliputi informasi historis terkait penjualan gorden yang berisi minggu penjualan, motif gorden, jumlah penjualan. Pengumpulan data dilakukan selama periode [1 Januari 2023 – 22 Maret 2025]. Random Forest merupakan metode yang sangat cocok untuk memprediksi penjualan ritel jangka pendek, terutama ketika data historis yang tersedia terbatas [12]. Data merupakan kumpulan informasi bersifat abstrak yang setelah dihimpun dan disusun dalam satu kesatuan dapat memberi pemahaman lebih detail dan menyeluruh mengenai objek atau fenomena tertentu [13]

2.2 Exploratory Data Analysis (EDA)

Langkah berikutnya adalah melakukan Exploratory Data Analysis (EDA) guna mendalami dan memahami pola serta karakteristik data secara lebih menyeluruh. Proses analisis memanfaatkan library visualisasi Python seperti Matplotlib dan Seaborn, serta Scikit-Learn untuk mendukung dokumentasi teknis dan proses seleksi fitur. Visualisasi data dalam EDA dapat dilakukan dengan berbagai metode, seperti grafik, scatter plot, histogram, serta statistik deskriptif yang melibatkan perhitungan rata-rata, median, dan standar deviasi, yang merupakan teknik yang umum digunakan [14].

2.3 Preprocessing Data

Pada tahap preprocessing data, langkah pertama yang dilakukan adalah pembersihan data dengan menangani nilai yang hilang (missing values). Selanjutnya, dilakukan transformasi data dengan mengubah format data ke dalam long format, di mana kolom minggu yang terpisah diubah menjadi satu variabel dengan nilai penjualan yang terkumpul dalam satu kolom.

Setelah itu, dilakukan Feature Engineering dengan membuat berbagai fitur tambahan yang relevan untuk pemodelan. Proses ini meliputi pembuatan fitur lag yang merepresentasikan jumlah penjualan pada minggu sebelumnya, rolling statistics yang mencerminkan rata-rata dan standar deviasi dari tiga minggu terakhir, fitur Tren juga dihitung untuk mengukur perubahan penjualan dari minggu sebelumnya, statistik agregat dihitung untuk setiap motif guna memberikan gambaran yang lebih lengkap mengenai perilaku penjualan per motif.

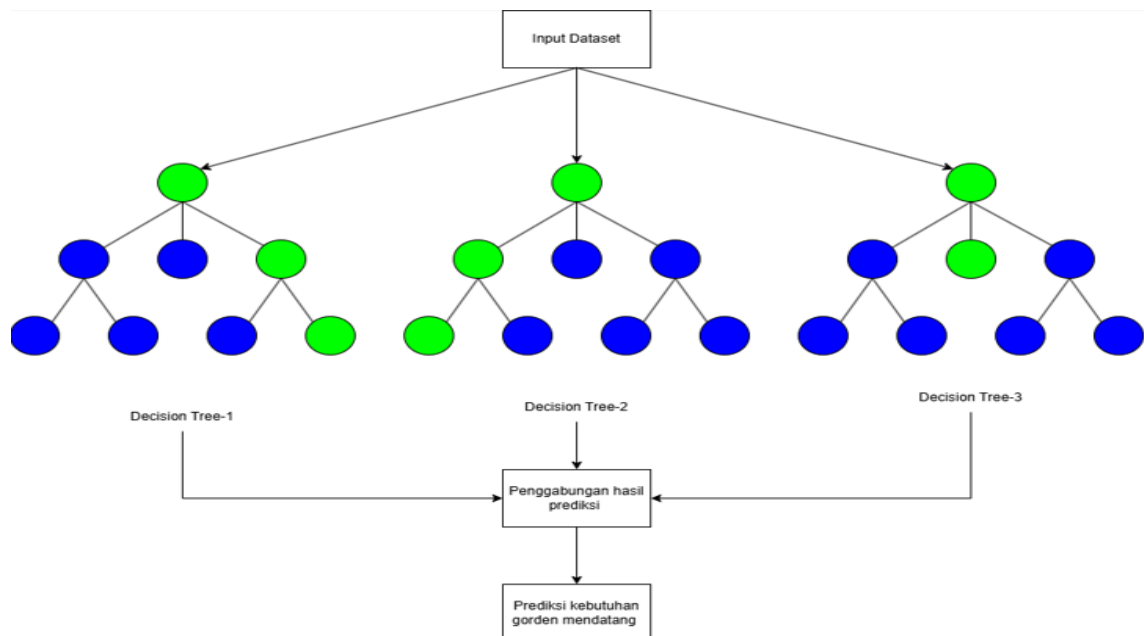
Fitur cyclical juga dibuat dengan menggunakan fungsi sin dan cos untuk merepresentasikan pola musiman berdasarkan minggu dalam setahun (52 minggu). Proses encoding untuk mengubah kategori motif menjadi representasi numerik. Proses ini memastikan data siap optimal sebelum diterapkan pada analisis maupun pemodelan berikutnya [15].

2.4 Data Splitting

Langkah berikutnya adalah data splitting, yaitu proses membagi dataset menjadi tiga bagian utama: 70% digunakan sebagai data latih, 15% sebagai data validasi, dan 15% digunakan untuk data uji. Data penjualan yang merupakan deret waktu (time series), pembagian data tidak boleh dilakukan secara acak karena dapat mengganggu urutan waktu yang sangat penting dalam analisis prediksi. Oleh karena itu, pembagian data dilakukan secara kronologis, di mana data uji diambil dari minggu-minggu terakhir dalam dataset. Data latih digunakan untuk melatih model agar dapat mengenali pola hubungan antara variabel input dan output. Data validasi digunakan untuk mengevaluasi kinerja model selama proses pelatihan, memungkinkan penyesuaian hyperparameter dan penghindaran overfitting, sehingga model dapat lebih generalisasi terhadap data yang belum terlihat. Data uji digunakan untuk menguji kemampuan model dalam memprediksi data baru yang belum pernah dilihat sebelumnya. Tujuan dari membagi dataset adalah untuk melatih model menggunakan data yang berbeda dari data yang digunakan untuk menguji kinerja model [16].

2.5 Modeling

Tahapan ini merupakan inti dari penelitian, yaitu membangun model prediksi menggunakan algoritma Random Forest. Metode ini adalah algoritma ensemble yang membuat beberapa pohon keputusan dalam proses pelatihan dan menggabungkan hasilnya untuk membuat prediksi akhir [17]. Metode ini banyak diterapkan dalam pembelajaran mesin untuk tugas-tugas seperti klasifikasi dan regresi [18]. Dalam penelitian ini algoritma Random Forest digunakan untuk memprediksi jumlah penjualan gorden mendatang berdasarkan data historis penjualan pada Tova Gorden, sekaligus untuk mengidentifikasi kategori gorden dengan tingkat penjualan tinggi dan rendah guna mendukung pengambilan keputusan stok yang lebih akurat. Pemilihan atribut terbaik untuk memisahkan node pada setiap pohon dilakukan dengan menggunakan Mean Squared Error (MSE) yang bertujuan untuk meminimalkan kesalahan prediksi dan menghasilkan pohon keputusan yang lebih akurat dan informatif.



Gambar 2. Diagram alur pembentukan model Random Forest

Pada tahap awal, dataset masukan diolah untuk membentuk beberapa pohon keputusan (Decision Tree-1, Decision Tree-2, dan Decision Tree-3), di mana masing-masing pohon menghasilkan prediksi secara mandiri. Selanjutnya, seluruh hasil prediksi tersebut digabungkan (penggabungan hasil prediksi) untuk memperoleh prediksi akhir kebutuhan gorden pada periode selanjutnya. Gambar 2 menampilkan diagram alur proses pembentukan model Random Forest yang digunakan dalam penelitian ini.

2.6 Evaluasi hasil prediksi akurasi dan Tingkat kesalahan

Evaluasi dilakukan untuk menilai performa model Random Forest dalam memprediksi jumlah penjualan gorden pada Tova Gorden dengan menggunakan data penjualan historis sebagai dasar analisis. Dalam penelitian ini digunakan tiga metrik evaluasi utama, yaitu Root Mean Squared Error (RMSE) merupakan indikator yang umum digunakan untuk mengevaluasi sejauh mana prediksi model mendekati nilai sebenarnya [19], Mean Absolute Percentage Error (MAPE) digunakan untuk menghitung tingkat kesalahan prediksi dalam bentuk persentase terhadap nilai aktual sehingga dapat memberikan gambaran seberapa besar perbedaan antara hasil prediksi dan data sebenarnya [20], Mean Absolute Error (MAE) digunakan untuk Menilai rata-rata selisih absolut antara hasil prediksi dan nilai yang sebenarnya [21], dan Koefisien Determinasi (R^2) digunakan untuk menilai sejauh mana variasi jumlah stok dapat dijelaskan oleh model berdasarkan data penjualan sebelumnya. Semakin mendekati nilai 1, maka semakin baik kemampuan model dalam menjelaskan variasi data yang ada, menunjukkan bahwa model tersebut lebih efektif dalam memprediksi perubahan stok berdasarkan pola penjualan yang teramati [22].

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

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (2)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (3)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (4)$$

Pada penelitian ini, terdapat beberapa variabel yang digunakan untuk mengevaluasi kinerja model prediksi. Variabel pertama adalah n , yang menunjukkan jumlah data yang dianalisis. Selanjutnya, terdapat y_i , yaitu nilai aktual pada data ke- i yang diamati dalam studi ini. Sebagai pembandingan, \hat{y}_i merupakan nilai prediksi yang dihasilkan untuk setiap data ke- i . Selain itu, \bar{y} merujuk pada rata-rata dari nilai aktual yang dihitung untuk keseluruhan data yang digunakan dalam analisis ini.

2.7 Deployment

Tahapan ini merupakan langkah penerapan model prediksi yang telah dikembangkan ke dalam sistem operasional Tova Gorden. Model prediksi diimplementasikan dan dijalankan melalui Google Colaboratory. Antarmuka pengguna dibuat dengan menggunakan framework Flask, yang memungkinkan pengembangan aplikasi web secara cepat dan ringan [23].

3. HASIL DAN PEMBAHASAN

Penelitian ini menggunakan algoritma Random Forest untuk memprediksi jumlah penjualan gorden secara mingguan di Tova Gorden. Sebelum proses pemodelan dilakukan, tahap persiapan data melalui rekayasa fitur (feature engineering) dilakukan untuk memperkaya dataset. Pada tahap ini, fitur-fitur baru yang relevan dan informatif ditambahkan untuk membantu model dalam memahami pola penjualan gorden. Beberapa fitur yang ditambahkan antara lain lag features, rolling statistics, Statistik agregat, fitur siklikal, dan encoding motif. Model dilatih menggunakan dataset total 7.772 sampel, menjadi 70% data latih, 15% data validasi dan 15% data, menghasilkan 5.443 sampel untuk data latih, 1.163 sampel untuk data validasi, dan 1.166 sampel untuk data uji.

3.1 Model Random Forest

Dalam penelitian ini, model Random Forest Regressor diterapkan untuk memprediksi jumlah penjualan gorden secara mingguan di Tova Gorden. Random Forest merupakan metode ensemble learning yang menggunakan sejumlah pohon keputusan (decision trees) untuk membuat prediksi. Setiap pohon keputusan dilatih pada subset data yang berbeda, dan prediksi akhir dihasilkan dengan mengambil rata-rata hasil dari semua pohon yang terlibat. Pendekatan ini efektif untuk mengurangi masalah overfitting, yang sering terjadi pada model berbasis pohon keputusan tunggal, sekaligus meningkatkan stabilitas dan akurasi prediksi. Penelitian ini menggunakan Random Forest Regressor dengan beberapa parameter utama, di antaranya :

- $n_estimators=300$: Menggunakan 300 pohon keputusan untuk meningkatkan ketepatan dan konsistensi prediksi.



- b. `max_depth=25`: Menetapkan kedalaman maksimum pohon keputusan untuk mencegah model menjadi terlalu kompleks dan mengalami overfitting.
- c. `min_samples_split=3`: Menentukan jumlah minimum sampel yang dibutuhkan untuk membagi suatu node pada pohon, yang mencegah pembagian node pada data yang sangat sedikit.
- d. `min_samples_leaf=2`: Menetapkan jumlah minimum sampel pada daun pohon keputusan untuk memastikan model tetap dapat memberikan prediksi yang stabil.
- e. `max_features='sqrt'`: Menggunakan subset acak dari fitur yang tersedia pada setiap pembagian pohon untuk mengurangi ketergantungan pada fitur tertentu dan meningkatkan kemampuan generalisasi model.
- f. `random_state=42`: Parameter ini memastikan hasil pelatihan dapat direplikasi, sehingga eksperimen yang dilakukan dapat menghasilkan output yang konsisten.
- g. `n_jobs=-1`: Memanfaatkan seluruh inti prosesor untuk mempercepat proses pelatihan model dengan menjalankan beberapa proses secara paralel.

Pemilihan parameter-parameter ini dilakukan berdasarkan eksperimen dan penyesuaian nilai secara bertahap, dengan menguji berbagai kombinasi parameter untuk menentukan mana yang memberikan hasil terbaik pada model. Proses ini melibatkan evaluasi model terhadap data yang ada, dan parameter yang dipilih merupakan hasil dari pencocokan dan penyesuaian yang dilakukan untuk mencapai performa model yang optimal.

3.2 Evaluasi Model

Beberapa metrik evaluasi digunakan untuk menilai kualitas prediksi model, yang diuji pada tiga jenis data, yaitu data latih, data validasi, dan data uji. Di antara metrik tersebut adalah RMSE, MAPE, MAE, dan R².

Tabel 1. Evaluasi Model Random Forest

Dataset	RMSE	MAPE (%)	MAE	R ² (%)
Latih	2.6215	12.90 %	1.1296	0.9783 %
Validasi	4.6639	25.54 %	2.4736	0.9372%
Uji	3.2345	26.63%	2.2940	0.9664 %

RMSE (Root Mean Squared Error) digunakan untuk mengukur rata-rata selisih kuadrat antara nilai aktual dan nilai prediksi. Semakin kecil nilai RMSE, semakin baik kualitas prediksi model. Pada data latih, RMSE adalah 2.6215, yang berarti rata-rata prediksi pada data latih memiliki kesalahan sekitar ±3 unit stok. Pada data validasi, RMSE adalah 4.6639, menunjukkan sedikit peningkatan kesalahan pada data validasi dengan rata-rata sekitar ±5 unit stok. Sementara pada data uji, RMSE adalah 3.2345, yang menunjukkan hasil yang sedikit lebih baik dibandingkan dengan data validasi dengan kesalahan prediksi sekitar ±3 unit stok.

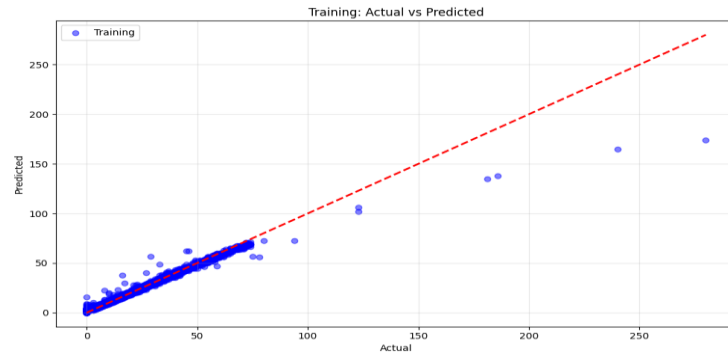
MAPE (Mean Absolute Percentage Error) digunakan untuk mengukur rata-rata persentase kesalahan antara nilai aktual dan prediksi. Pada data latih, MAPE sebesar 12.90%, yang berarti bahwa prediksi model rata-rata menyimpang sekitar 12.90% dari nilai aktual. Pada data Validasi, MAPE meningkat menjadi 25.54%, menunjukkan adanya peningkatan kesalahan prediksi. Pada data uji, MAPE meningkat lagi menjadi 26.63%, yang menunjukkan kesalahan prediksi sedikit lebih besar dibandingkan dengan dataset latih.

MAE (Mean Absolute Error) digunakan untuk mengukur rata-rata kesalahan absolut antara prediksi dan nilai aktual. Nilai MAE yang lebih rendah menunjukkan prediksi yang lebih tepat. Pada data latih, MAE adalah 1.1296, yang berarti rata-rata kesalahan prediksi adalah sekitar ±1 unit stok. Pada data validasi, MAE adalah 2.4736, yang menunjukkan sedikit peningkatan pada kesalahan prediksi dengan rata-rata sekitar ±2 unit stok. Sementara pada data uji, MAE adalah 2.2940, yang berarti rata-rata kesalahan prediksi juga sekitar ±2 unit stok.

R² (Koefisien Determinasi) digunakan untuk mengukur seberapa besar variasi dalam data yang dapat dijelaskan oleh model. Semakin mendekati 100%, semakin baik model dalam menjelaskan variasi data. Pada data latih, R² sebesar 0.9783, yang menunjukkan bahwa model dapat menjelaskan 97.83% variasi dalam data latih dengan sangat baik. Pada data validasi, R² sebesar 0.9372, yang berarti model masih dapat menjelaskan 93.72% variasi dalam data validasi dengan baik. Sementara itu, pada data uji, R² sebesar 0.9664, yang menunjukkan bahwa model juga sangat baik dalam menjelaskan 96.64% variasi data pada set uji.

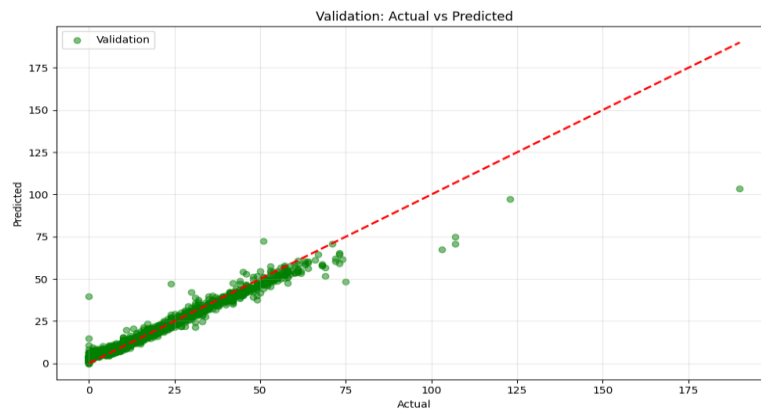
Secara keseluruhan, model ini berhasil menunjukkan performa yang sangat baik dengan nilai R² yang konsisten tinggi di setiap dataset, yang menunjukkan bahwa model memiliki kemampuan yang kuat dalam menjelaskan variasi data. Namun, sedikit peningkatan kesalahan pada data validasi dan uji mengindikasikan adanya kemungkinan overfitting pada model, yang perlu diperhatikan untuk meningkatkan generalisasi model ke data baru.

Berdasarkan Gambar 3, grafik ini menggambarkan perbandingan antara nilai aktual dan prediksi pada data pelatihan. Titik-titik biru menunjukkan hasil prediksi model Random Forest, sementara garis merah putus-putus menunjukkan garis ideal di mana nilai prediksi seharusnya berimpit dengan nilai aktual. Sebagian besar titik berada dekat dengan garis merah, menandakan bahwa model memiliki akurasi prediksi yang sangat baik. Meskipun demikian, ada beberapa titik yang sedikit menyimpang, mengindikasikan adanya kesalahan prediksi pada beberapa observasi. Secara keseluruhan, model menunjukkan kemampuan yang sangat baik dalam memprediksi jumlah penjualan pada data pelatihan.



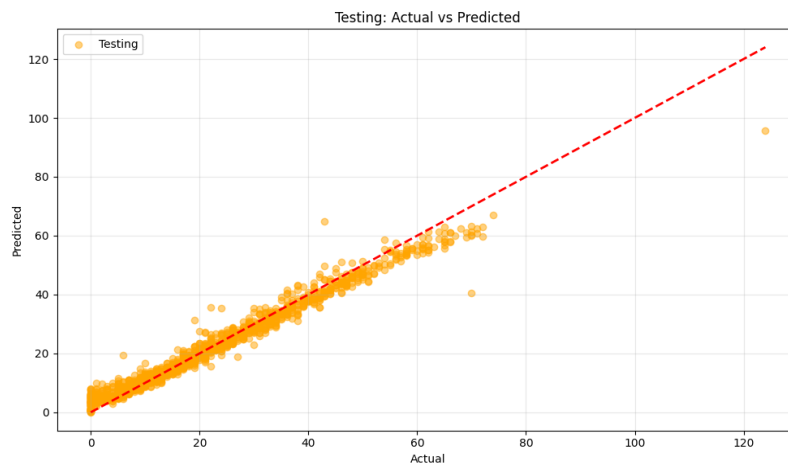
Gambar 3. Visualisasi Perbandingan Aktual vs prediksi (Latih)

Gambar 4 menunjukkan perbandingan antara nilai aktual dan prediksi pada data validasi. Titik hijau mewakili hasil prediksi model, sementara garis merah putus-putus menunjukkan garis ideal di mana nilai aktual dan prediksi seharusnya berimpit. Sebagian besar titik berada dekat dengan garis tersebut, menandakan model mempertahankan akurasi yang baik pada data validasi. Meskipun ada beberapa titik yang menyimpang, terutama pada nilai aktual yang lebih tinggi, secara keseluruhan model menunjukkan kinerja yang stabil dan dapat menggeneralisasi dengan baik pada data validasi.



Gambar 4. Visualisasi Perbandingan Aktual vs prediksi (Validasi)

Gambar 5. menunjukkan perbandingan antara nilai aktual dan prediksi pada data uji. Titik oranye mewakili hasil prediksi model, sementara garis merah putus-putus menggambarkan garis ideal yang menunjukkan kesempurnaan prediksi. Sebagian besar titik berada dekat dengan garis tersebut, menunjukkan akurasi prediksi yang sangat baik. Meskipun ada sedikit penyimpangan pada nilai prediksi yang lebih rendah, secara keseluruhan, model menunjukkan performa yang sangat baik dalam memprediksi jumlah penjualan pada data yang belum dilihat sebelumnya.



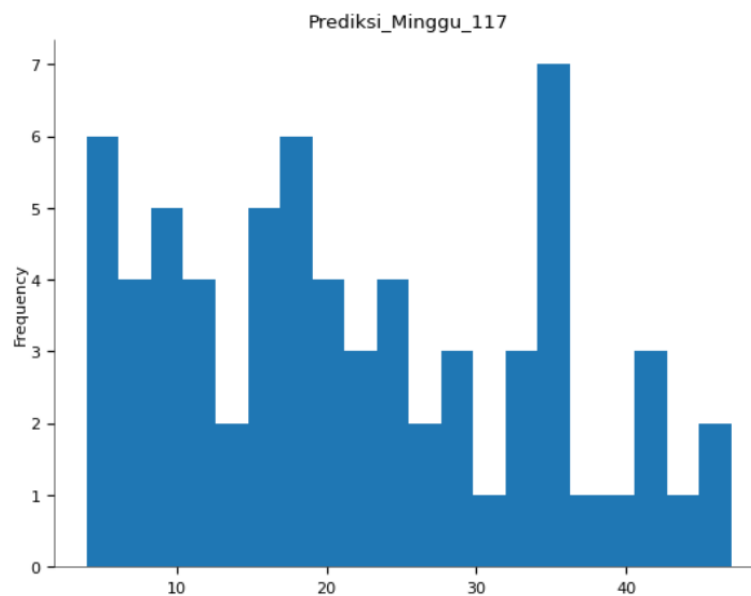
Gambar 5. Visualisasi Perbandingan Aktual vs prediksi (uji)

3.3 Prediksi Model

Setelah model Random Forest dilatih dan dievaluasi, langkah berikutnya adalah menggunakan model untuk memprediksi jumlah penjualan gorden mingguan di Tova Gorden.

3.3.1 Prediksi jumlah penjualan gorden

Prediksi dilakukan berdasarkan fitur-fitur yang telah dipersiapkan selama tahap feature engineering, seperti motif gorden, jumlah penjualan minggu sebelumnya, dan faktor musiman. Model kemudian memproses data yang ada dan memberikan estimasi jumlah penjualan gorden untuk setiap minggu. Hasil prediksi yang dihasilkan akan menunjukkan jumlah unit stok gorden yang diperkirakan akan terjual. Dengan informasi ini, tim pembelian di Tova Gorden dapat lebih efektif dalam merencanakan dan mengelola stok, memastikan ketersediaan produk yang sesuai dengan permintaan, serta mengurangi kemungkinan kelebihan atau kekurangan stok yang dapat merugikan perusahaan. Melalui proses ini, model tidak hanya membantu dalam perencanaan stok yang lebih efisien, tetapi juga memberikan wawasan berharga bagi perusahaan dalam menghadapi fluktuasi permintaan dan tren pasar yang dinamis. Dengan demikian, model prediksi penjualan gorden ini diharapkan dapat meningkatkan efisiensi operasional dan meminimalisir risiko ketidakpastian dalam proses manajemen stok di Tova Gorden. Gambar 6. menunjukkan distribusi prediksi jumlah penjualan gorden untuk minggu ke 117. Terlihat adanya konsentrasi pada rentang nilai tertentu, dengan beberapa titik memiliki frekuensi yang lebih besar dibandingkan titik lainnya. Contohnya, nilai antara 10 hingga 20 unit muncul lebih sering, menunjukkan bahwa sebagian besar produk diperkirakan terjual pada rentang tersebut. Namun, terdapat juga prediksi dengan angka penjualan yang lebih tinggi atau lebih rendah, meskipun frekuensinya lebih sedikit.



Gambar 6. Distribusi Prediksi Jumlah Penjualan

3.3.2 Kategori Jumlah Penjualan

Agar pengambilan keputusan lebih mudah, hasil prediksi penjualan gorden dibagi menjadi dua kategori. Penjualan dianggap tinggi jika nilai prediksi ≥ 30 , dan rendah jika nilai prediksi kurang dari 30. Sebagai ilustrasi, jika hasil prediksi jumlah penjualan gorden untuk minggu tertentu adalah 39 unit, maka produk tersebut dikategorikan sebagai penjualan tinggi karena nilainya lebih besar dari atau sama dengan 30. Sebaliknya, jika prediksi menunjukkan 4 unit, produk tersebut akan dikategorikan sebagai penjualan rendah, karena nilai prediksinya kurang dari 30. Ambang batas 30 untuk kategori penjualan tinggi dan rendah ditentukan berdasarkan kebijakan pemilik Tova Gorden, yang mempertimbangkan efisiensi pengelolaan stok dan kebutuhan pasar berdasarkan data historis. Pembagian kategori ini memudahkan pengambilan keputusan terkait pengadaan stok, sehingga produk dengan prediksi penjualan tinggi dapat diprioritaskan dalam perencanaan persediaan. Pada tabel 2 ditampilkan hasil prediksi dan kategori jumlah penjualan.

Tabel 2. Contoh Hasil Prediksi dan Kategori Jumlah penjualan

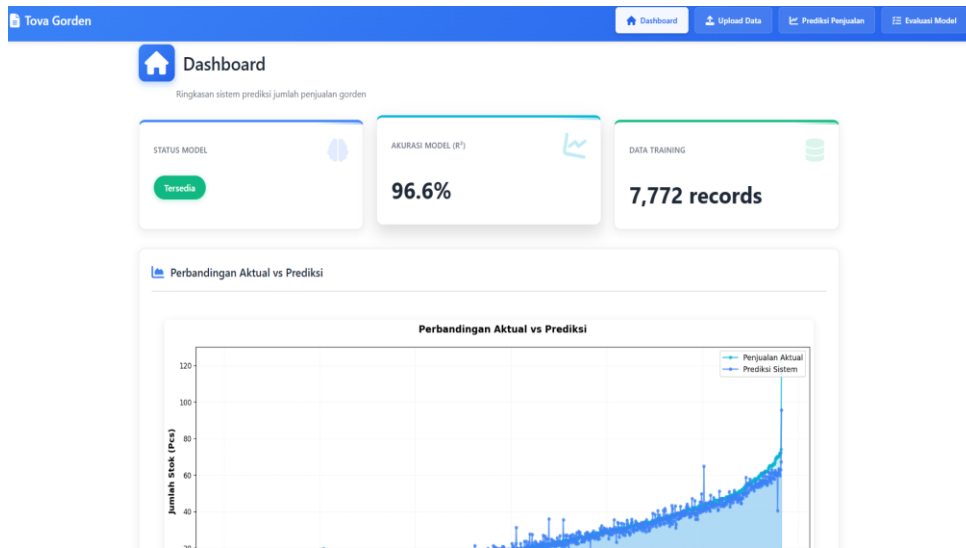
Motif	Prediksi Jumlah	Kategori
abu_muda	4	Rendah
Green_tulip	39	Tinggi
gucci	20	Rendah

3.4 Implementasi Sistem Prediksi

Model Random Forest diterapkan dalam sistem prediksi melalui pengembangan antarmuka visual berbasis web, menggunakan framework Flask. Flask sebagai framework yang digunakan memastikan bahwa aplikasi berjalan dengan efisien, ringan, dan dapat diakses dengan mudah, baik untuk penggunaan internal maupun pengembangan

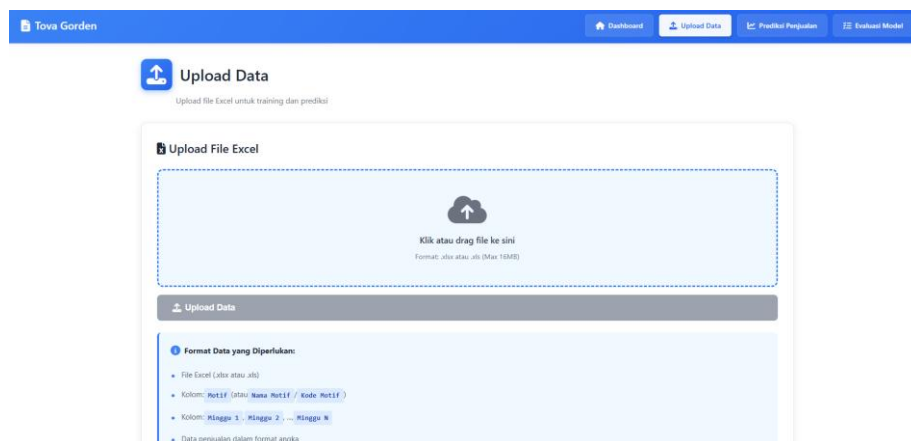
lebih lanjut. Sistem ini menyediakan antarmuka pengguna berbasis web yang interaktif, memungkinkan pengguna untuk dengan mudah mengunggah data penjualan, menampilkan hasil prediksi jumlah dan kategori penjualan, serta menampilkan hasil evaluasi model. Dengan antarmuka yang ramah pengguna, sistem ini memungkinkan pengguna untuk melakukan prediksi dan evaluasi model dengan efisien dan mudah.

Gambar 7. menunjukkan tampilan halaman Dashboard dari sistem prediksi jumlah penjualan gorden yang telah diimplementasikan. Halaman ini menyajikan ringkasan status model, akurasi model, dan jumlah data pelatihan yang digunakan. Pengguna dapat melihat secara langsung hasil perbandingan antara nilai aktual dan prediksi jumlah penjualan, yang membantu memvisualisasikan kinerja model dalam memprediksi penjualan. Dengan desain yang sederhana dan intuitif, halaman Dashboard ini memudahkan pengguna untuk mengelola dan memonitor hasil prediksi secara efektif.



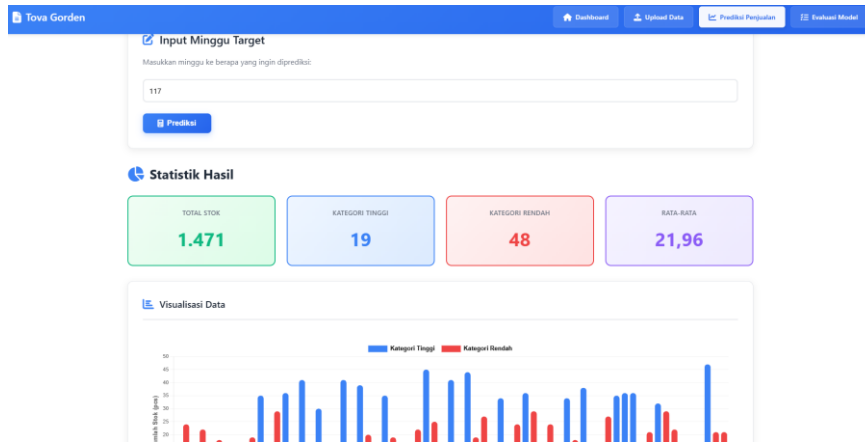
Gambar 7. Tampilan Halaman Dashboard

Gambar 8 menunjukkan tampilan halaman Upload Data dalam sistem prediksi. Pada halaman ini, pengguna dapat mengunggah file Excel yang berisi data penjualan yang akan digunakan untuk pelatihan model atau untuk prediksi. Sistem ini menyediakan area untuk menyeret atau memilih file secara langsung, dengan format file yang jelas, yaitu .xlsx dengan ukuran maksimal 16MB. Selain itu, di bagian bawah halaman, terdapat instruksi mengenai format data yang harus diikuti, termasuk kolom yang diperlukan seperti kolom Motif, Nama Motif, dan data penjualan mingguan. Hal ini memastikan bahwa data yang diunggah memiliki struktur yang sesuai untuk diproses dalam model prediksi.



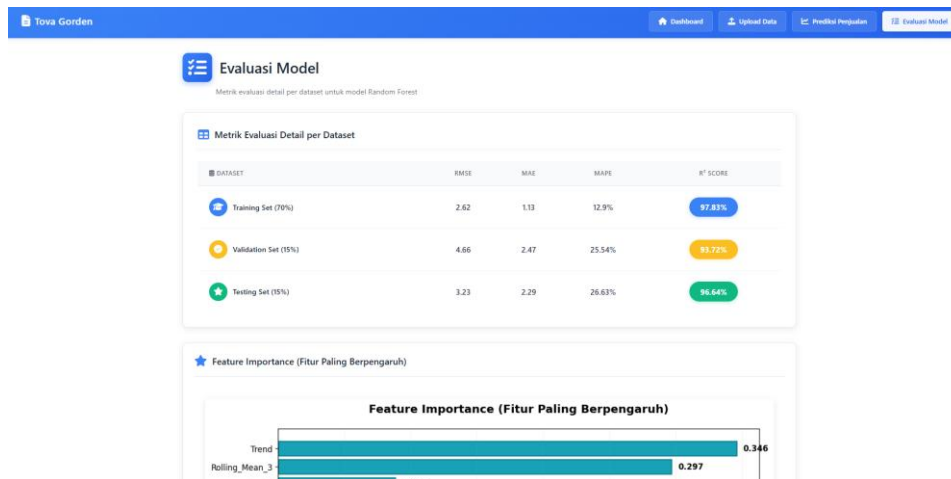
Gambar 8. Tampilan Halaman Upload Data

Pada Gambar 9 ditampilkan antarmuka halaman prediksi penjualan. Halaman ini memungkinkan pengguna untuk memasukkan minggu target yang ingin diprediksi dan langsung melihat hasil prediksi jumlah stok yang dibutuhkan. Setelah pengguna memilih minggu target, sistem menampilkan statistik hasil prediksi yang mencakup total stok, kategori tinggi dan rendah, serta rata-rata stok yang diprediksi. Selain itu, hasil prediksi juga disajikan dalam bentuk visualisasi yang memudahkan pengguna untuk memahami distribusi stok dan kategori prediksi. Sebagai tambahan, detail hasil prediksi untuk masing-masing motif gorden juga ditampilkan, di mana setiap motif dikategorikan berdasarkan jumlah stok yang diprediksi, apakah termasuk kategori tinggi atau rendah. Dengan informasi ini, tim pengelola dapat lebih mudah menentukan prioritas pengadaan stok yang efisien.



Gambar 9. Tampilan Halaman Prediksi Penjualan

Gambar 10 menampilkan halaman evaluasi model yang digunakan untuk mengevaluasi performa model Random Forest berdasarkan data latih, validasi, dan uji. Halaman ini menyajikan metrik evaluasi untuk setiap dataset, meliputi RMSE, MAE, MAPE, dan R^2 (koefisien determinasi). Nilai-nilai metrik menunjukkan kinerja model pada setiap set data yang digunakan. Sebagai contoh, pada data latih, model memperoleh nilai R^2 sebesar 97.83%, yang menunjukkan bahwa model sangat baik dalam menjelaskan variabilitas data latih. Selanjutnya, tampilan juga menyertakan grafik Feature Importance yang menunjukkan fitur-fitur yang paling berpengaruh dalam model. Dalam hal ini, fitur Rolling_Mean_3 memiliki tingkat pengaruh tertinggi.



Gambar 10. Tampilan Halaman Evaluasi Model

4.KESIMPULAN

Penelitian ini telah berhasil mengembangkan dan mengimplementasikan model prediksi jumlah penjualan gorden menggunakan algoritma Random Forest. Proses pelatihan model dilakukan dengan menggunakan data historis penjualan gorden dari Tova Gorden, yang telah diproses dengan tahap feature engineering untuk meningkatkan akurasi prediksi. Berdasarkan evaluasi model menggunakan data latih, validasi, dan uji, model menunjukkan performa yang sangat baik, dengan nilai R^2 mencapai 97.83% pada data latih, 93.72% pada data validasi, dan 96.64% pada data uji. Harus diperhatikan bahwa model ini hanya dilatih dan diuji menggunakan data dari Tova Gorden. Dengan demikian, meskipun model ini menunjukkan kinerja yang sangat baik dalam pengujian internal, belum dapat dipastikan apakah model ini juga dapat diterapkan di toko lain. Selain itu, hasil penelitian menunjukkan bahwa Rolling Mean (tren jangka pendek) merupakan fitur yang paling berpengaruh dalam meningkatkan akurasi prediksi. Fitur ini terbukti kunci dalam model, karena mengidentifikasi pola penjualan yang berulang dalam periode pendek, sehingga memberikan kontribusi signifikan terhadap ketepatan prediksi. Keterbatasan dari sistem prediksi ini terletak pada penggunaan data yang terbatas hanya pada Tova Gorden. Model ini dikembangkan dan diuji menggunakan data historis penjualan dari satu sumber, sehingga belum ada validasi apakah model ini dapat diadaptasi atau diterapkan dengan hasil yang sama baiknya pada toko atau pasar lainnya. Selain itu, meskipun fitur-fitur yang digunakan dalam rekayasa data sudah cukup lengkap, faktor-faktor eksternal seperti perubahan tren pasar, promosi, atau kebijakan perusahaan lain yang berbeda dapat memengaruhi akurasi prediksi model. Oleh karena itu, penerapan model di luar konteks Tova Gorden memerlukan evaluasi ulang dan



penyesuaian terhadap variabel atau fitur yang relevan, agar sistem dapat memberikan prediksi yang akurat dan dapat diandalkan di lingkungan yang berbeda. Sebagai saran untuk penelitian selanjutnya, variabel eksternal, seperti jadwal promosi media sosial atau perubahan tren pasar, dapat diintegrasikan ke dalam model untuk meningkatkan akurasi prediksi. Variabel tersebut diharapkan dapat memberikan wawasan lebih mendalam mengenai faktor-faktor yang memengaruhi permintaan, sehingga model dapat lebih adaptif terhadap perubahan yang terjadi di pasar. Secara keseluruhan, sistem prediksi yang dibangun dapat memberikan kontribusi positif dalam perencanaan dan manajemen stok gorden di Tova Gorden, serta meningkatkan efisiensi operasional dengan meminimalkan risiko kekurangan atau kelebihan stok.

REFERENCES

- [1] L. P. Arista and Y. S. Nugroho, "Sistem Informasi Pencatatan Transaksi Penjualan Dan Pembelian Produk Berbasis Website Di Toko Sembako Putrasena Sukoharjo," *J. Inform. Polinema*, vol. 9, no. Sistem Informasi Pencatatan Transaksi Penjualan Dan Pembelian Produk Berbasis Website Di Toko Sembako Putrasena Sukoharjo, pp. 397–404, 2023, doi: 10.33795/jip.v9i4.1347.
- [2] B. A. Sekti, A. P. Gusti, and N. Erzed, "Perancangan Sistem Informasi Stok Barang berbasis Web dengan Metode FIFO," *J. Teknol. Inform. dan Komput.*, vol. 10, no. 2, pp. 506–518, 2024, doi: 10.37012/jtik.v10i2.2253.
- [3] A. Afandi, I. N. Farida, and U. Mahdiyah, "Penerapan Algoritma Apriori Dan Metode Moving Average Untuk Prediksi Stok Barang," *Pros. SEMNAS INOTEK (Seminar Nas. Inov. Teknol.*, vol. 6, no. 2, pp. 421–426, 2022, doi: <https://doi.org/10.29407/inotek.v6i2.2624>.
- [4] P. K. Andini, A. Indrawan, and E. Martaseli, "Analisis Sistem Pengendalian Internal Atas Persediaan Barang Dagang (Sparepart) Dalam Upaya Mempertahankan Laba Pada PT. Selamat Lestari Mandiri," *Akunt. 45*, vol. 5, no. 2, pp. 643–656, 2024, doi: 10.30640/akuntansi45.v5i2.3386.
- [5] S. Ristayanti, I. Shalshabilla, R. A. Rachmawan, and D. E. Susilo, "Analisis Peran Sistem Informasi Dalam Meningkatkan Efektivitas Pengendalian Persediaan Barang Dagang Pada PT. Mayora Indah Tbk.," *JASIKA (Jurnal Sist. Inf. Akuntansi)*, vol. 04, no. 02, pp. 74–82, 2024, doi: <https://doi.org/10.31294/jasika.v4i2.3757>.
- [6] G. N. Ayuni and D. Fitriana, "Penerapan Metode Regresi Linear Untuk Prediksi Penjualan Properti pada PT XYZ," *J. Telemat.*, vol. 14, no. 2, pp. 79–86, 2020, doi: 10.61769/telematika.v14i2.321.
- [7] J. Luo and Y. Shen, "Optimization of Random Forest Algorithm and Research on the Effectiveness of its Application in Stock Index Forecasting," *J. Comb. Math. Comb. Comput.*, vol. 127b, no. April 2024, pp. 9557–9571, 2025, doi: 10.61091/jcmcc.127b-523.
- [8] D. Manurung, B. Zealtiel, and A. H. Lubis, "Prediksi Produksi Tanaman Padi di Indonesia dengan Menggunakan Algoritma Random Forest Regressor," *J. Comput. Informatics Res.*, vol. 4, no. 3, pp. 337–345, 2025, doi: 10.47065/comforch.v4i3.2125.
- [9] R. Hidayat et al., "Implementasi Algoritma Random Forest Regression Untuk Memprediksi Penjualan Produksi di Supermarket," *Simkom*, vol. 10, no. 1, pp. 101–109, 2025, doi: 10.51717/simkom.v10i1.703.
- [10] R. Verdianto, D. Hartanti, and E. Purwanto, "Pengembangan Aplikasi Point of Sales untuk Prediksi Penjualan Harian Usaha Minuman Menggunakan Algoritma Random Forest Regression," *Infotek J. Inform. dan Teknol.*, vol. 8, no. 1, pp. 128–139, 2025, doi: 10.29408/jit.v8i1.28386.
- [11] A. Ridwan, U. Muzakir, and S. Nurhidayati, "Optimizing E-commerce Inventory to prevent Stock Outs using the Random Forest Algorithm Approach," *Int. J. Softw. Eng. Comput. Sci.*, vol. 4, no. 1, pp. 107–120, 2024, doi: 10.35870/ijsecs.v4i1.2326.
- [12] M. Ismail, H. M. Tukur, and M. Friday, "Sales Prediction using Ensemble Machine Learning Model," *Int. J. Sci. Res. Mod. Technol.*, vol. 4, no. 3, pp. 24–35, 2025, doi: <https://doi.org/10.38124/ijrsmt.v4i3.350>.
- [13] M. A. M. Setiawan, K. Kusri, and A. D. Hartono, "Menggunakan Metode Machine Learning Untuk Memprediksi Nilai Mahasiswa Dengan Model Prediksi Multiclass," *J. Inform. J. Pengemb. IT*, vol. 10, no. 1, pp. 190–204, 2025, doi: 10.30591/jpit.v10i1.8334.
- [14] N. Nurdiansyah, F. S. Febriyan, Z. Gesit, and D. Amanta, "Mental Health Analysis to Prevent Mental Disorders in Students Using The K-Nearest Neighbor (K-NN) Algorithm and Random Forest Algorithm Analisis Kesehatan Mental untuk Mencegah Gangguan Mental pada Mahasiswa Menggunakan Algoritma K-Nearest Neighbor (K)," *Inst. Ris. dan Publ. Indones.*, vol. 5, no. January, pp. 1–9, 2025, doi: <https://doi.org/10.57152/malcom.v5i1.1537>.
- [15] A. Roni and Y. Amri, "Sentimen Analisis Aplikasi Posaja Pada Google Playstore Untuk Peningkatan Pospay Superapp Menggunakan Support Vector Meachine," *J. Tek. Inform.*, vol. 16, no. 2, pp. 1–7, 2024, [Online]. Available: <https://ejurnal.ulbi.ac.id/index.php/informatika/article/view/3533>
- [16] H. Tantyoko, D. K. Sari, and A. R. Wijaya, "PREDIKSI POTENSIAL GEMPA BUMI INDONESIA MENGGUNAKAN METODE RANDOM FOREST DAN FEATURE SELECTION," *Indones. J. Inf. Syst.*, vol. 6, pp. 83–89, 2023, doi: <https://doi.org/10.36080/idealis.v6i2.3036>.
- [17] H. Imaduddin, W. Widayat, and F. Y. A'la, "Classification of Diabetes using Machine Learning," 2021 Int. Conf. Comput. Perform. Eval. ComPE 2021, pp. 185–189, 2021, doi: 10.1109/ComPE53109.2021.9751955.
- [18] H. P. Khandagale, R. Patil, S. Patil, and D. Bhosale, "Predicting Stock Prices With Machine Learning Using Comparative Analysis of Random Forest Algorithm," *Int. J. Eng. Appl. Sci. Technol.*, vol. 8, no. 6, pp. 60–68, 2023, doi: 10.33564/ijeast.2023.v08i06.008.
- [19] Z. Sun, "The Approaches to Predict Stock Prices by Random Forest," *Appl. Comput. Eng.*, vol. 96, no. 1, pp. 162–168, 2024, doi: 10.54254/2755-2721/96/20241333.
- [20] R. W. Wilda, M. A. Sukmarini, and A. C. Mahar, "Prediksi Harga Saham PT. Unilever Indonesia TBK Dengan Metode Regresi Linier Sederhana," *Balanc. Media Inf. Akunt. dan Keuang.*, vol. 16, no. 2, pp. 76–81, 2024, doi: 10.52300/blnc.v16i2.14249.
- [21] N. Yulis, M. A. Anhar, and A. S. Rombe, "Analisis Perbandingan Peramalan Penggunaan Bahan Baku Menggunakan



- Metode Weighted Moving Average (WMA) dan Evaluasi dengan Mean Absolute Error (MAE),” Pros. Semin. Ilm. Sist. Inf. DAN Teknol. Inf., vol. XIV, no. 1, pp. 57–64, 2025, doi: <https://doi.org/10.36774/sisiti.v14i1.1675>.
- [22] F. Putra, Susanti, Herwin, Kh. Andesa, and Mardainis, “Prediksi Nilai Redaman Jaringan Fiber Optik untuk Menilai Kinerja Jaringan Menggunakan Random Forest Regression,” *Indones. J. Comput. Sci.*, vol. x, no. x, pp. 3347–3361, 2025, doi: <https://doi.org/10.33022/ijcs.v14i2.4796> Prediksi.
- [23] C. Wijayanto and Y. A. Susetyo, “Implementasi flask framework pada pembangunan aplikasi sistem informasi helpdesk (sih),” *JUPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.*, vol. 07, no. September, pp. 858–868, 2022, doi: [10.29100/jipi.v7i3.3161](https://doi.org/10.29100/jipi.v7i3.3161).