

Prediksi Total Penjualan Gorden Berdasarkan Data Historis Bulanan Menggunakan Random Forest Regressor

Aliya Aszava^{1*}, Aulia Happy Salma², Muhammad Azka Latif³, Muhammad Fajar Maulana⁴,
 Muhammad Arifin⁵

^{1,2,3,4,5}Program Studi Sistem Informasi Fakultas Teknik, Universitas Muria Kudus
 E-mail: 202353044@std.umk.ac.id^{1*}, 202353018@std.umk.ac.id², 202353012@std.umk.ac.id³,
 202353015@std.umk.ac.id⁴, arifin.m@umk.ac.id⁵

INFO ARTIKEL

Sejarah Artikel

Diterima : 02/05/2026

Direvisi : 11/05/2026

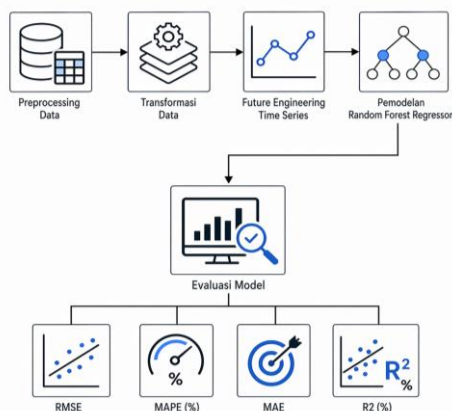
Diterbitkan : 01/06/2026

*Corresponding author

202353044@std.umk.ac.id

DOI: 10.70247/jumistik.v5i1.292

GRAPHICAL ABSTRACT



ABSTRACT

Forecasting monthly sales is important to support stock planning and operational decisions, especially in curtain businesses with fluctuating demand. This study aims to predict total monthly curtain sales at Azka Gorden using Random Forest Regressor based on historical data. The study applied a quantitative approach with data mining stages consisting of data collection, preprocessing, data transformation, time series feature engineering, modeling, and evaluation. The initial dataset consisted of 63 monthly observations from January 2021 to March 2026. After temporal, seasonal, lag, and moving average features were created, 51 observations were ready for modeling and were split chronologically into training, validation, and testing sets. Model performance was evaluated using RMSE, MAPE, MAE, and R^2 . The results show R^2 values of 98.56% on training data, 94.87% on validation data, and 98.18% on testing data, with low prediction error. These results indicate that Random Forest Regressor can support more accurate monthly sales prediction and stock planning.

Keywords: sales forecasting, Random Forest Regressor, time series, curtain sales, inventory planning

ABSTRAK

Prediksi penjualan bulanan penting untuk mendukung perencanaan stok dan keputusan operasional, terutama pada usaha gorden yang memiliki permintaan fluktuatif. Penelitian ini bertujuan memprediksi total item penjualan gorden bulanan di Azka Gorden menggunakan Random Forest Regressor berbasis data historis. Penelitian menerapkan pendekatan kuantitatif dengan tahapan *data mining* yang meliputi pengumpulan data, *preprocessing*, transformasi data, *feature engineering time series*, pemodelan, dan evaluasi. *Dataset* awal terdiri atas 63 observasi bulanan dari Januari 2021 sampai Maret 2026. Setelah pembentukan fitur temporal, musiman, *lag*, dan *moving average*, diperoleh 51 observasi yang siap dimodelkan dan dibagi secara kronologis menjadi data latih, validasi, dan uji. Kinerja model dievaluasi menggunakan RMSE, MAPE, MAE, dan R^2 . Hasil penelitian menunjukkan nilai R^2 sebesar 98,56% pada data latih, 94,87% pada data validasi, dan 98,18% pada data uji, dengan kesalahan prediksi yang rendah. Hasil ini menunjukkan model efektif untuk mendukung prediksi penjualan dan perencanaan stok yang lebih tepat.

Kata kunci: prediksi penjualan, Random Forest Regressor, deret waktu, penjualan gorden, perencanaan persediaan

© 2026 Penerbit STMIK Amika Soppeng. All rights reserved

PENDAHULUAN

Perkembangan teknologi informasi dan meningkatnya persaingan usaha mendorong pelaku bisnis untuk menggunakan data sebagai dasar pengambilan keputusan. Dalam kegiatan penjualan, keputusan mengenai persediaan, pengadaan barang, dan strategi pemasaran tidak cukup hanya bertumpu pada intuisi. Data penjualan historis dapat diolah untuk membaca pola permintaan, memperkirakan kebutuhan periode berikutnya, dan mengurangi ketidaktepatan keputusan operasional. Oleh karena itu, *forecasting* berbasis data menjadi pendekatan penting bagi usaha ritel yang menghadapi perubahan permintaan secara berkala [1].

Kebutuhan terhadap prediksi penjualan semakin penting pada produk rumah tangga dan interior seperti gorden [2]. Permintaan produk ini dapat berubah karena kebutuhan dekorasi rumah, pola belanja musiman, tren motif, pilihan bahan, serta daya beli konsumen. Ketika perubahan permintaan tidak dianalisis dengan baik, pelaku usaha berisiko mengalami kelebihan stok yang menambah biaya penyimpanan atau kekurangan stok yang menyebabkan peluang penjualan hilang dan menurunkan kepuasan pelanggan [3].

Dalam konteks penjualan gorden, perencanaan stok memerlukan pemahaman terhadap pola bulanan dan kecenderungan berulang. Total penjualan dapat naik atau turun mengikuti periode tertentu sehingga pendekatan prediksi perlu mempertimbangkan aspek waktu, bukan hanya menghitung rata-rata penjualan. Dengan memanfaatkan data bulanan, pelaku usaha dapat memperoleh gambaran yang lebih objektif mengenai arah permintaan dan menyiapkan keputusan persediaan secara lebih terukur.

Dalam praktiknya, banyak usaha kecil dan menengah masih melakukan perencanaan penjualan secara manual berdasarkan pengalaman atau perkiraan subjektif [2]. Cara ini sederhana, tetapi kurang memadai ketika pola penjualan semakin kompleks dan jumlah periode pengamatan bertambah. Penelitian terdahulu menunjukkan bahwa ketidakpastian penjualan dapat berdampak pada pengelolaan stok dan perencanaan produksi [1]. Pada sektor ritel, proses manual juga berisiko memicu ketidakseimbangan persediaan dan menurunkan kinerja penjualan [4].

Random Forest Regressor dipilih karena mampu membangun banyak pohon keputusan dan menggabungkan hasilnya untuk memperoleh prediksi numerik yang lebih stabil [5]. Karakteristik ini sesuai dengan permasalahan prediksi penjualan gorden, karena data penjualan bulanan dapat memuat pola nonlinier, pengaruh periode sebelumnya, serta gejala musiman. Dibandingkan pendekatan manual, model regresi berbasis *Random Forest* dapat mempelajari hubungan antarfitur secara lebih sistematis dan mengurangi risiko keputusan stok yang hanya didasarkan pada perkiraan subjektif [6], [7].

Permasalahan yang melatarbelakangi penelitian ini dapat dirumuskan dalam beberapa hal. Pertama, diperlukan model prediksi yang mampu memperkirakan total penjualan gorden secara objektif berdasarkan *dataset* bulanan. Kedua, perencanaan persediaan masih membutuhkan dasar analisis yang lebih terukur agar risiko *overstock* dan *understock* dapat dikurangi. Ketiga, pola penjualan bulanan perlu dimodelkan dengan fitur deret waktu, seperti *lag*, *moving average*, dan fitur musiman, sehingga hubungan antara histori penjualan dan nilai target dapat dikenali. Keempat, perlu diketahui sejauh mana *Random Forest Regressor* mampu menghasilkan prediksi yang layak digunakan sebagai dasar pendukung keputusan.

Berdasarkan uraian tersebut, penelitian ini dilakukan untuk memprediksi total penjualan gorden bulanan menggunakan *Random Forest Regressor*. Kegiatan penelitian diarahkan untuk menyiapkan *dataset* bulanan, membentuk fitur deret waktu, membangun model regresi, dan mengevaluasi hasil prediksi menggunakan metrik RMSE, MAPE, MAE, dan R^2 . Hasil yang diharapkan adalah model prediksi yang dapat memberikan estimasi penjualan secara lebih terukur, sehingga dapat menjadi dasar perencanaan stok dan referensi bagi penelitian sejenis pada bidang prediksi penjualan berbasis data bulanan.

TINJAUAN PUSTAKA

Wijanarko et al. (2026) meneliti prediksi jumlah penjualan gorden pada Tova Gorden yang mengalami masalah ketidakefisienan stok akibat permintaan yang fluktuatif. Permasalahan utama dalam penelitian ini adalah sulitnya menentukan jumlah persediaan yang tepat ketika permintaan berubah dari waktu ke waktu. Untuk mengatasi hal tersebut, peneliti menggunakan *Random Forest Regressor* pada data historis penjualan mingguan. Prosesnya dilakukan melalui tahap *preprocessing*, *feature engineering*, pembagian data secara kronologis, serta evaluasi model menggunakan RMSE, MAPE, MAE, dan R^2 . Hasil penelitian menunjukkan performa model yang sangat baik dengan nilai R^2 sebesar 97,83% pada data latih, 93,72% pada data validasi, dan 96,64% pada data uji. Temuan ini memperlihatkan bahwa *Random Forest* efektif untuk *forecasting* jumlah penjualan berbasis data historis, khususnya ketika tujuan penelitian bukan sekadar mengelompokkan kategori penjualan, melainkan memperkirakan nilai penjualan secara kuantitatif.

Farida et al. (2025) meneliti usaha rumah makan yang mengalami fluktuasi penjualan akibat promosi, tren musiman, hari libur, dan acara khusus. Penelitian ini menerapkan *Random Forest* pada data penjualan rumah makan setelah melalui tahap *preprocessing* dan evaluasi menggunakan *accuracy*, *precision*, *recall*, dan *F1-score*. Hasil penelitian menunjukkan akurasi sebesar 97 persen pada data uji, sehingga model dinilai efektif untuk mendukung pengelolaan operasional dan strategi pemasaran. Hasil serupa juga ditunjukkan oleh Ramadhan et al. (2025) pada rumah

makan seafood. Penelitian tersebut memakai data primer selama satu tahun dan pendekatan KDD untuk memprediksi tren penjualan. Hasil evaluasi menunjukkan *accuracy* sebesar 96,25 persen, *precision* 0,9633, *recall* 0,9624, dan *F1-score* 0,9627. Kedua studi ini menegaskan bahwa *Random Forest* mampu memberi prediksi penjualan yang baik pada sektor kuliner.

Efendi et al. (2024) yang menerapkan *Random Forest* untuk prediksi penjualan sekaligus pengelolaan persediaan produk Bolen Crispy. Penelitian ini menggunakan data penjualan selama satu tahun, dengan pembagian 80% data pelatihan dan 20% data pengujian. Evaluasi model menunjukkan akurasi 0,85, MAE 1,2, dan RMSE 1,8. Hasil akhirnya tidak hanya menampilkan prediksi penjualan untuk satu minggu ke depan, tetapi juga menunjukkan bahwa model membantu pengelolaan stok agar lebih efisien dan mengurangi risiko *overstock* maupun *understock*. Penelitian ini menegaskan bahwa *Random Forest* tidak hanya berguna untuk memprediksi penjualan, tetapi juga relevan dalam mendukung keputusan persediaan pada usaha kecil menengah.

Nusri et al. (2026) menunjukkan penerapan *Random Forest* pada konteks yang berbeda, yaitu prediksi pemesanan tiket online. Penelitian ini berangkat dari masalah fluktuasi permintaan tiket yang menyulitkan penyedia layanan dalam mengelola stok dan strategi penjualan. Untuk menyelesaikannya, peneliti mengembangkan sistem prediksi berbasis *Random Forest* dengan memanfaatkan 5.000 transaksi tiket yang dipilih secara acak. Variabel yang digunakan meliputi harga tiket, waktu pemesanan, jenis acara, lokasi acara, dan jumlah tiket yang tersedia. Data diproses melalui pembersihan, normalisasi, pemisahan data latih dan data uji, lalu model dibandingkan dengan regresi linear dan *decision tree* tunggal. Hasil penelitian menunjukkan bahwa *Random Forest* memberikan performa terbaik dengan MAE 0,142, RMSE 0,185, dan R^2 0,892. Nilai tersebut lebih baik dibandingkan regresi linear maupun *decision tree* tunggal, dan perbedaannya juga signifikan secara statistik dengan $p\text{-value} < 0,001$. Temuan ini memperlihatkan bahwa *Random Forest* mampu memberikan prediksi yang akurat untuk mendukung optimasi penjualan dan manajemen inventaris, meskipun objek penelitiannya berada di luar bidang kuliner.

Dalam pemodelan prediksi berbasis data bulanan, fitur deret waktu menjadi komponen penting karena model *machine learning* tidak secara otomatis memahami urutan periode. Data transaksi perlu ditransformasikan menjadi deret waktu agar setiap observasi merepresentasikan satu periode pengamatan yang terukur [10]. Fitur *lag* digunakan untuk membawa informasi nilai penjualan pada periode sebelumnya, sedangkan *moving average* membantu menangkap kecenderungan jangka pendek atau menengah. Fitur musiman seperti *month_sin* dan *month_cos* digunakan untuk merepresentasikan siklus bulan dalam satu tahun,

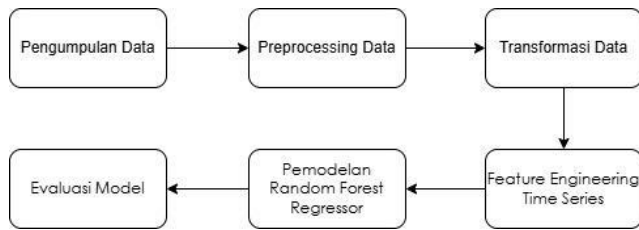
sehingga model dapat membaca pola berulang tanpa memperlakukan bulan sebagai angka linear biasa. Penggunaan fitur temporal tersebut relevan untuk *Random Forest Regressor* karena algoritma ini membaca pola dari variabel masukan yang telah dibentuk secara eksplisit [11]. Studi prediksi berbasis *machine learning* lain juga menunjukkan bahwa evaluasi model regresi dapat diperluas melalui perbandingan beberapa algoritma ketika penelitian berikutnya membutuhkan peningkatan performa [12].

Berdasarkan seluruh penelitian tersebut, *Random Forest* menunjukkan kinerja yang baik untuk prediksi penjualan dan kebutuhan operasional pada berbagai objek. Namun, sebagian penelitian terdahulu masih berfokus pada klasifikasi tren atau kategori penjualan, sedangkan penelitian lain menekankan integrasi sistem persediaan jangka pendek. Berbeda dengan studi tersebut, penelitian ini menekankan prediksi kuantitatif total penjualan gorden bulanan dengan memanfaatkan kombinasi fitur temporal, fitur musiman, *lag*, dan *moving average*. *Research gap* penelitian ini terletak pada kebutuhan model prediksi yang tidak hanya membaca tren umum, tetapi juga mengolah pola deret waktu bulanan sebagai dasar estimasi jumlah penjualan. Dengan demikian, *state of the art* penelitian ini adalah penerapan *Random Forest Regressor* pada *dataset* penjualan gorden bulanan melalui *feature engineering* deret waktu yang lebih terstruktur, sehingga hasil prediksi lebih langsung dapat digunakan untuk mendukung perencanaan stok dan keputusan operasional.

METODOLOGI PENELITIAN

Tahapan Penelitian

Penelitian ini menggunakan metode kuantitatif dengan pendekatan *data mining* untuk membangun model prediksi penjualan berbasis data historis. Pendekatan ini dipilih karena penelitian berfokus pada pengolahan data penjualan secara sistematis agar diperoleh model yang mampu memprediksi jumlah penjualan pada periode berikutnya. Model yang digunakan adalah *Random Forest Regressor* karena variabel target pada penelitian ini berbentuk numerik, yaitu total item penjualan gorden bulanan. Tahapan penelitian meliputi pengumpulan data, *preprocessing* data, transformasi data, *feature engineering time series*, pemodelan *Random Forest Regressor*, dan evaluasi model. Urutan tahapan tersebut dirancang agar alur pengolahan data berjalan konsisten mulai dari data mentah sampai evaluasi performa model pada data uji. Tahapan penelitian tersebut dapat disajikan pada Gambar 1.



Gambar 1. Tahapan Penelitian
Sumber: Hasil olah data penulis

Pengumpulan Data

Tahap pengumpulan data dilakukan dengan menggunakan data penjualan historis yang tersimpan dalam file Excel. Data yang digunakan merupakan data historis penjualan bulanan Azka Gorden yang diperoleh dari pencatatan penjualan internal. Data awal memuat 63 observasi dari Januari 2021 sampai Maret 2026. Variabel awal terdiri atas Periode, Tahun, Bulan_Ke, Kuartal, Trend Index, Gorden Dilan, Gorden Keraton, Gorden Mahira, Gorden Melamin, dan Total Gorden. Variabel Gorden Dilan, Gorden Keraton, Gorden Mahira, dan Gorden Melamin merepresentasikan jumlah penjualan per kelompok produk, sedangkan Total Gorden digunakan sebagai variabel target dalam pemodelan. Data tersebut selanjutnya digunakan sebagai dasar pembentukan fitur deret waktu dan pembangunan model prediksi total penjualan gorden bulanan.

Analisis Data

Preprocessing Data

Pada tahap *preprocessing*, data diperiksa untuk memastikan konsistensi format, urutan waktu, dan kelengkapan nilai. Kolom Periode dikonversi ke format tanggal, data diurutkan secara kronologis, dan baris ringkasan total dikeluarkan dari proses analisis karena tidak merepresentasikan satu periode bulanan. Selain itu, variabel numerik disesuaikan agar dapat digunakan dalam pembentukan fitur deret waktu. Tahap ini bertujuan menghasilkan data yang bersih, konsisten, dan siap digunakan pada proses berikutnya [10].

Transformasi Data

Setelah data melalui tahap *preprocessing*, data ditransformasikan ke dalam bentuk *dataset* analitis yang siap digunakan untuk pemodelan. Pada penelitian ini, variabel target yang digunakan adalah Total Gorden, yaitu jumlah penjualan total item pada setiap periode bulanan. Transformasi data dilakukan dengan memilih variabel yang relevan untuk analisis, sehingga setiap baris data merepresentasikan satu periode pengamatan dengan satu nilai target penjualan. Tahap ini bertujuan mengubah data operasional menjadi data analitis yang sesuai untuk proses *machine learning*.

Feature Engineering Time Series

Tahap berikutnya adalah pembentukan fitur *time series*. Tahap ini dilakukan untuk menerjemahkan pola

historis penjualan menjadi variabel prediktif yang dapat dibaca oleh model Random Forest Regressor [11]. Fitur yang digunakan meliputi fitur waktu, yaitu Tahun, Bulan_Ke, Kuartal, dan Trend Index. Selain itu, digunakan pula fitur musiman berbentuk *month_sin* dan *month_cos* untuk menangkap pola berulang antarbulan dalam satu siklus tahunan. Penelitian ini juga membentuk fitur *lag* pada 1, 2, 3, 6, dan 12 bulan sebelumnya agar model mampu membaca pengaruh penjualan masa lalu terhadap penjualan saat ini. Selain fitur *lag*, dibentuk pula *moving average* pada jendela 3 bulan dan 6 bulan untuk menangkap kecenderungan penjualan jangka pendek dan menengah. Setelah seluruh fitur terbentuk, baris data yang belum memiliki histori cukup untuk pembentukan fitur tersebut tidak digunakan dalam proses pemodelan. Tahap ini bertujuan memperkaya informasi masukan sehingga model dapat mengenali pola tren, pola musiman, dan dinamika historis penjualan dengan lebih baik.

Pemodelan Random Forest Regressor

Data yang telah melalui *preprocessing*, transformasi, dan *feature engineering* selanjutnya digunakan untuk membangun model prediksi menggunakan Random Forest Regressor [13]. Algoritma ini dipilih karena mampu menangani hubungan nonlinier antarvariabel dan sesuai untuk prediksi data numerik. Pada tahap ini, data dibagi secara kronologis menjadi data latih, data validasi, dan data uji. Pembagian kronologis dilakukan agar urutan waktu tetap terjaga dan tidak terjadi kebocoran informasi dari periode masa depan ke periode sebelumnya. Pengaturan dasar model yang digunakan pada eksperimen adalah *n_estimators=500*, *max_depth=10*, *min_samples_leaf=2*, *max_features="sqrt"*, dan *random_state=42*. Pengaturan tersebut dicantumkan agar eksperimen dapat diuji ulang dengan konfigurasi yang sama.

$$\hat{y}(x) = \frac{1}{B} \sum_{b=1}^B T_b(x) \quad (1)$$

Pada Persamaan (1), $\hat{y}(x)$ menyatakan nilai prediksi akhir model, *B* menyatakan jumlah pohon keputusan yang digunakan, dan $T_b(x)$ menyatakan hasil prediksi dari pohon keputusan ke-*b* terhadap input *x*. Dengan mekanisme ini, Random Forest tidak bergantung pada satu pohon saja, melainkan pada gabungan banyak pohon, sehingga prediksi yang dihasilkan cenderung lebih stabil.

Evaluasi Model

Kinerja model dievaluasi menggunakan metrik regresi, yaitu Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), dan koefisien determinasi (R^2). MAE digunakan untuk mengukur rata-rata selisih absolut antara nilai aktual dan nilai prediksi. RMSE digunakan untuk menilai besar kesalahan prediksi dengan penekanan lebih kuat pada error yang tinggi. MAPE digunakan untuk melihat persentase deviasi hasil

prediksi terhadap nilai aktual. Sementara itu, R^2 digunakan untuk menilai kemampuan model dalam menjelaskan variasi data. Evaluasi dilakukan pada data latih, validasi, dan uji agar performa model dapat diamati secara menyeluruh.

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

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (3)$$

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

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (5)$$

Pada Persamaan (2), (3), (4), dan (5), y_i adalah nilai aktual, $\hat{y}_i$ adalah nilai prediksi, $\bar{y}$ adalah rata-rata nilai aktual, dan n adalah jumlah observasi pada data yang dievaluasi. RMSE menunjukkan besar kesalahan prediksi dengan penekanan lebih kuat pada kesalahan yang besar, MAPE menyatakan persentase kesalahan relatif terhadap nilai aktual, MAE menggambarkan rata-rata selisih absolut, dan R^2 menunjukkan kemampuan model dalam menjelaskan variasi data. Hasil perhitungan metrik evaluasi disajikan pada bagian Hasil dan Pembahasan.

HASIL DAN PEMBAHASAN

Penelitian ini menggunakan algoritma Random Forest Regressor untuk memprediksi jumlah total item penjualan gorden secara bulanan di Azka Gorden. Sebelum proses pemodelan dilakukan, tahap pembentukan fitur *time series* melalui *Feature engineering Time series* untuk merepresentasikan pola historis penjualan. *Dataset* awal terdiri atas 63 observasi bulanan pada periode Januari 2021 sampai Maret 2026. Setelah pembentukan fitur *lag* dan *moving average*, jumlah data siap model menjadi 51 observasi, yang selanjutnya dibagi secara kronologis menjadi 35 data latih, 7 data validasi, dan 9 data uji. Pembagian kronologis dilakukan agar urutan waktu tetap terjaga dan evaluasi model lebih sesuai dengan karakteristik data deret waktu.

Pengumpulan Data

Hasil pengumpulan data menunjukkan bahwa data penjualan berasal dari rekap bulanan pada periode Januari 2021 sampai Maret 2026. Setelah data dibaca dan diurutkan, diperoleh 63 observasi bulanan yang memuat informasi periode, variabel waktu, dan total penjualan gorden pada setiap bulan. Variabel target yang digunakan pada penelitian ini adalah Total Gorden, sehingga setiap observasi merepresentasikan jumlah total item gorden yang terjual pada satu periode bulanan. Struktur data ini sesuai dengan tujuan penelitian, yaitu membangun model prediksi penjualan total berbasis data historis bulanan.

Data awal menunjukkan pola penjualan yang berulang dan relatif halus antarbulan. Nilai Total

Gorden berada pada rentang 4.060 sampai 4.420 item per bulan. Keberadaan variabel waktu seperti tahun, bulan, kuartal, dan Trend Index memberikan dasar untuk membentuk fitur temporal. Dengan struktur tersebut, *dataset* layak digunakan untuk menguji kemampuan model regresi dalam mengenali pola tren dan musiman penjualan.

Preprocessing Data

Hasil *preprocessing* menunjukkan bahwa kolom periode berhasil dikonversi ke format tanggal, sehingga seluruh observasi dapat disusun secara kronologis dari periode paling awal hingga periode paling akhir. Selain itu, variabel numerik telah berada pada format yang konsisten, sehingga data siap digunakan dalam pembentukan fitur *time series* dan proses pelatihan model. Hasil tersebut menunjukkan bahwa data berhasil dipertahankan sebagai deret waktu bulanan yang utuh dan layak untuk dianalisis lebih lanjut.

Transformasi Data

Hasil transformasi menunjukkan bahwa data telah tersusun dalam bentuk *dataset* analitis yang siap digunakan untuk pemodelan [14]. Pada tahap ini, Total Gorden ditetapkan sebagai variabel target, sehingga setiap baris data merepresentasikan satu periode bulanan dengan satu nilai target penjualan. Struktur ini membuat hubungan antara variabel prediktor dan variabel target menjadi lebih jelas, sehingga data siap digunakan pada tahap pembentukan fitur dan pemodelan.

Feature Engineering Time Series

Tahap *feature engineering* menghasilkan serangkaian fitur *time series* yang digunakan untuk menangkap pola historis penjualan. Fitur tersebut meliputi variabel waktu, yaitu Tahun, Bulan_Ke, Kuartal, dan Trend Index. Selain itu, dibentuk pula fitur musiman *month_sin* dan *month_cos* untuk mewakili siklus penjualan bulanan dalam satu tahun. Di samping fitur waktu, penelitian ini membentuk fitur historis berupa *lag_1*, *lag_2*, *lag_3*, *lag_6*, dan *lag_12*, serta *moving average ma_3* dan *ma_6*. Kombinasi fitur ini dirancang agar model dapat mengenali pengaruh penjualan bulan sebelumnya, pola musiman tahunan, dan kecenderungan penjualan jangka pendek.

Karena fitur *lag* dan *moving average* membutuhkan histori penjualan periode sebelumnya, maka tidak semua observasi awal dapat langsung digunakan. Setelah baris yang belum memiliki histori cukup dihapus, jumlah data siap model menjadi 51 observasi. Artinya, terdapat pengurangan observasi dari data awal akibat kebutuhan pembentukan *lag* sampai 12 bulan. Hal ini wajar dalam analisis *time series*, karena model baru dapat dibangun ketika histori minimum telah tersedia. Hasil *feature engineering* ini menunjukkan bahwa pembentukan variabel prediktif berhasil dilakukan dan data siap dibagi menjadi data latih, data validasi, dan data uji. Pembagian data penelitian disajikan pada Tabel 1.

Tabel 1. Pembagian data penelitian

Dataset	Jumlah Sampel	Rentang Periode
Data awal	63	2021-01 s.d. 2026-03
Data siap model	51	2022-01 s.d. 2026-03
Data latih	35	2022-01 s.d. 2024-11
Data validasi	7	2024-12 s.d. 2025-06
Data uji	9	2025-07 s.d. 2026-03

Sumber: Hasil olah data penulis

Tabel 1 menunjukkan perubahan jumlah data dari tahap awal sampai pembagian data akhir. Pengurangan dari 63 menjadi 51 observasi terjadi setelah pembentukan fitur *lag* dan *moving average*. Tabel tersebut juga menunjukkan bahwa pembagian data dilakukan secara kronologis dengan komposisi yang mendekati proporsi 70%, 15%, dan 15%. Dengan pembagian tersebut, model diuji pada data yang berada setelah data pelatihan sehingga hasil evaluasi lebih merepresentasikan kondisi prediksi periode berikutnya.

Pemodelan Random Forest Regressor

Hasil pemodelan menunjukkan bahwa Random Forest Regressor berhasil dilatih menggunakan 35 observasi pada data latih dengan 13 fitur prediktor. Model kemudian menghasilkan prediksi untuk data latih, validasi, dan uji menggunakan konfigurasi yang sama seperti dijelaskan pada bagian metodologi. Dengan skema tersebut, hasil yang diperoleh dapat dibandingkan antar-split tanpa mengubah struktur model.

Pada data latih dan data uji, pola prediksi yang dihasilkan model terlihat dekat dengan nilai aktual. Pada data validasi terdapat peningkatan galat, tetapi nilai evaluasi masih berada pada rentang yang baik. Kondisi ini mengindikasikan bahwa fitur temporal, fitur musiman, *lag*, dan *moving average* mampu membantu model membaca dinamika penjualan bulanan [15].

Evaluasi Model

Hasil evaluasi model *Random Forest Regressor* pada data latih, validasi, dan uji disajikan pada Tabel 2.

Tabel 2. Evaluasi model *Random Forest*

Dataset	RMSE	MAPE (%)	MAE	R²
Latih	12.2949	0.20 %	8.5492	0.9856
Validasi	16.1164	0.29 %	12.0340	0.9487
Uji	13.2207	0.22 %	9.1753	0.9818

Sumber: Hasil olah data penulis

Tabel 2 menunjukkan bahwa model memiliki kinerja yang sangat baik pada seluruh split data. Pada data latih, model menghasilkan RMSE sebesar 12,2949, MAPE sebesar 0,20%, MAE sebesar 8,5492, dan R²

sebesar 0,9856. Pada data validasi, model menghasilkan RMSE sebesar 16,1164, MAPE sebesar 0,29%, MAE sebesar 12,0340, dan R² sebesar 0,9487. Pada data uji, model memperoleh RMSE sebesar 13,2207, MAPE sebesar 0,22%, MAE sebesar 9,1753, dan R² sebesar 0,9818. Nilai R² yang mendekati 1 menunjukkan bahwa model mampu menjelaskan sebagian besar variasi data penjualan. Di sisi lain, nilai RMSE dan MAE yang relatif kecil menunjukkan bahwa selisih antara nilai aktual dan prediksi masih berada dalam rentang yang rendah.

Jika dilihat secara komparatif, performa terbaik terdapat pada data latih dengan nilai R² tertinggi dan galat paling rendah. Hal ini wajar karena model memang belajar langsung dari data latih. Pada data validasi, nilai RMSE dan MAE meningkat, tetapi peningkatan tersebut masih dalam batas yang wajar dan tidak menunjukkan penurunan kinerja yang ekstrem. Sementara itu, pada data uji, nilai R² kembali sangat tinggi dan nilai galat cenderung menurun dibandingkan data validasi. Kondisi ini menunjukkan bahwa model memiliki kemampuan generalisasi yang baik pada data yang tidak digunakan saat pelatihan.

Visualisasi Aktual dan Prediksi

Perbandingan antara data aktual dan hasil prediksi pada data latih dapat dilihat pada Gambar 2.

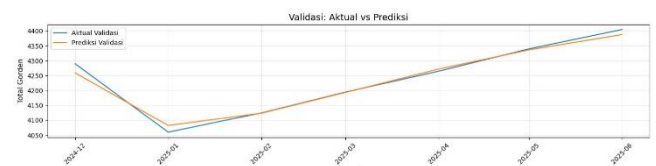


Gambar 2. Perbandingan Aktual dan Prediksi pada Data Latih

Sumber: Hasil olah data penulis

Gambar 2 memperlihatkan bahwa garis prediksi pada data latih sangat dekat dengan garis aktual. Pola kenaikan, penurunan, dan titik puncak penjualan pada sebagian besar bulan berhasil diikuti oleh model. Kedekatan dua garis ini sejalan dengan nilai R² yang tinggi pada data latih. Hasil tersebut menunjukkan bahwa model mampu mempelajari pola historis Total Gorden dengan baik selama proses pelatihan.

Perbandingan antara data aktual dan hasil prediksi pada data validasi dapat dilihat pada Gambar 3.

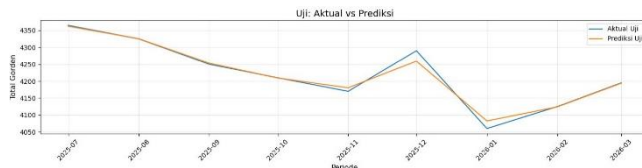


Gambar 3. Perbandingan Aktual dan Prediksi pada Data Validasi

Sumber: Hasil olah data penulis

Gambar 3 menunjukkan bahwa pada data validasi, garis prediksi tetap mengikuti arah perubahan data aktual. Walaupun terdapat sedikit perbedaan pada titik awal dan titik akhir periode validasi, secara umum model masih dapat merepresentasikan kecenderungan penjualan secara tepat. Hal ini memperlihatkan bahwa model tidak hanya cocok pada data pelatihan, tetapi juga cukup baik ketika dihadapkan pada data baru di luar proses pembelajaran.

Perbandingan antara data aktual dan hasil prediksi pada data uji dapat dilihat pada Gambar 4.



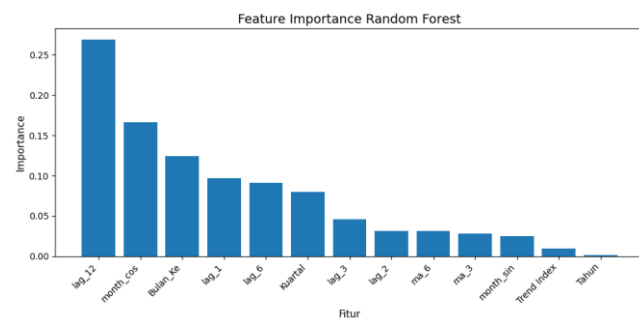
Gambar 4. Perbandingan Aktual dan Prediksi pada Data Uji

Sumber: Hasil olah data penulis

Gambar 4 memperlihatkan bahwa pola prediksi pada data uji masih konsisten mengikuti bentuk data aktual. Perbedaan terbesar terlihat ketika terjadi kenaikan pada Desember 2025 dan penurunan pada Januari 2026, namun secara umum model tetap mampu mendekati nilai aktual pada sebagian besar periode uji. Kedekatan garis aktual dan prediksi pada data uji menguatkan hasil evaluasi kuantitatif bahwa model memiliki kemampuan generalisasi yang baik terhadap data yang belum pernah dilihat sebelumnya.

Analisis Feature Importance

Kontribusi masing-masing fitur dalam pembentukan model *Random Forest* dapat dilihat pada Gambar 5.



Gambar 5. Feature importance Random Forest

Sumber: Hasil olah data penulis

Gambar 5 menunjukkan bahwa fitur *lag_12* memiliki nilai importance paling tinggi dibandingkan fitur lainnya. Setelah itu, fitur *month_cos*, *Bulan_Ke*, *Kuartal*, *lag_6*, dan *lag_1* juga memberikan kontribusi yang dominan dalam model. Sementara itu, *Trend Index* dan *Tahun* memiliki nilai importance yang relatif kecil. Hasil ini menunjukkan bahwa model lebih banyak bergantung pada pola musiman tahunan dan histori penjualan beberapa periode sebelumnya

daripada sekadar kecenderungan tren linear jangka panjang.

Pembahasan

Hasil evaluasi menunjukkan bahwa *Random Forest Regressor* mampu memprediksi total item penjualan gorden bulanan dengan performa yang sangat baik pada data latih, validasi, dan uji. Nilai R^2 yang tinggi disertai nilai RMSE, MAE, dan MAPE yang rendah menunjukkan bahwa model mampu menjelaskan variasi data penjualan secara kuat dan menghasilkan prediksi yang dekat dengan nilai aktual. Hasil ini memperlihatkan bahwa pendekatan berbasis data historis bulanan efektif untuk digunakan pada prediksi penjualan gorden di Azka Gorden.

Berdasarkan analisis *feature importance*, fitur *lag_12* menjadi variabel yang paling dominan, diikuti *month_cos*, *Bulan_Ke*, *Kuartal*, *lag_6*, dan *lag_1*. Temuan ini menunjukkan bahwa pola musiman tahunan dan histori penjualan periode sebelumnya memiliki peran utama dalam pembentukan prediksi. Dengan kata lain, model tidak hanya membaca kecenderungan jangka pendek, tetapi juga menangkap pola berulang yang muncul pada bulan yang sama di tahun sebelumnya. Hal ini sejalan dengan karakter data penjualan gorden yang cenderung dipengaruhi oleh siklus waktu tertentu.

Dari sisi praktis, model yang dibangun berpotensi digunakan sebagai alat bantu dalam perencanaan stok dan pengambilan keputusan operasional di Azka Gorden. Prediksi yang stabil dapat membantu pelaku usaha mengurangi risiko kekurangan maupun kelebihan stok [16]. Meskipun demikian, model ini masih bergantung pada histori data internal dan belum memasukkan faktor eksternal seperti promosi, hari besar, atau perubahan tren pasar [17]. Karena itu, pengembangan variabel pada penelitian berikutnya masih diperlukan agar model dapat menangkap dinamika penjualan secara lebih luas.

KESIMPULAN DAN SARAN

Penelitian ini berhasil membangun model prediksi total item penjualan gorden bulanan di Azka Gorden menggunakan *Random Forest Regressor* berbasis data historis. Hasil penelitian menunjukkan bahwa model memiliki kinerja yang sangat baik, dengan nilai R^2 sebesar 98,56% pada data latih, 94,87% pada data validasi, dan 98,18% pada data uji. Nilai galat model juga relatif rendah, dengan RMSE sebesar 12,2949 pada data latih, 16,1164 pada data validasi, dan 13,2207 pada data uji, sedangkan MAPE berada pada kisaran 0,20% sampai 0,29%. Nilai MAE juga relatif rendah, yaitu 8,5492 pada data latih, 12,0340 pada data validasi, dan 9,1753 pada data uji. Hasil perbandingan antara data aktual dan prediksi menunjukkan bahwa model mampu mengikuti pola perubahan penjualan dengan baik pada seluruh pembagian data. Selain itu, analisis *feature importance* menunjukkan bahwa *lag_12*, *month_cos*, *Bulan_Ke*, *Kuartal*, *lag_6*, dan *lag_1* menjadi fitur yang

paling berpengaruh, sehingga dapat disimpulkan bahwa pola musiman tahunan dan histori penjualan periode sebelumnya memegang peranan penting dalam prediksi Total Gorden. Hasil tersebut menunjukkan bahwa tujuan penelitian untuk membangun model prediksi penjualan bulanan berbasis *Random Forest Regressor* telah tercapai.

Penelitian selanjutnya disarankan menambahkan variabel eksternal yang berpotensi memengaruhi penjualan, seperti promosi, hari besar, tren permintaan pasar, dan faktor musiman lain di luar kalender dasar agar model mampu menangkap perubahan penjualan secara lebih luas. Selain itu, penelitian berikutnya dapat membandingkan *Random Forest Regressor* dengan algoritma lain, seperti XGBoost, LightGBM, atau LSTM, untuk melihat kemungkinan peningkatan performa model. Penambahan periode data historis juga penting agar model memiliki informasi yang lebih kaya dalam membaca pola jangka panjang. Pengembangan lebih lanjut juga dapat diarahkan pada integrasi model ke dalam sistem prediksi yang terhubung dengan *dashboard* penjualan agar hasil prediksi dapat dimanfaatkan langsung dalam pengambilan keputusan operasional di Azka Gorden.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Azka Gorden yang telah memberikan dukungan dalam penelitian ini melalui penyediaan data penjualan historis sebagai sumber utama analisis. Ucapan terima kasih juga disampaikan kepada semua pihak yang telah membantu proses penyusunan data, pengolahan data, dan pengembangan model prediksi sehingga penelitian ini dapat diselesaikan dengan baik.

DAFTAR PUSTAKA

- [1] E. S. Barus and D. Darmanto, "Implementasi Metode Random Forest Untuk Memprediksi Penjualan Produk," *Jurnal Teknik Informasi dan Komputer (Tekinkom)*, vol. 7, no. 2, p. 591, Dec. 2024, doi: 10.37600/tekinkom.v7i2.1510.
- [2] Y. M. Nurak, S. W. Iriananda, and F. Marisa, "Prediksi Penjualan Warung Kopi Ol Menggunakan Metode Random Forest dan XGBoost," *Jurnal Mahasiswa Teknik Informatika*, vol. 9, no. 4, pp. 5796–5802, Aug. 2025, doi: <https://doi.org/10.36040/jati.v9i4.13922>.
- [3] M. Syahrul Efendi et al., "Penerapan Algoritma Random Forest Untuk Prediksi Penjualan Dan Sistem Persediaan Produk," *Media Online*, vol. 5, no. 1, p. 20, Sep. 2024, doi: 10.30865/resolusi.v5i1.2149.
- [4] S. Falah, A. T. Jaka, and N. Q. Nada, "Penerapan Algoritma Regresi Linier Dan Random Forest Untukmemprediksi Penjualan Di Uchi Parfume," *Jurnal Informatika Teknologi dan Sains (JINTEKS)*, vol. 7, no. 3, pp. 1317–26, Aug. 2025, Accessed: May 10, 2026. [Online]. Available: <https://garuda.kemdiktisaintek.go.id/documents/detail/5376726>
- [5] E. P. Febtiawan, L. A. S. I. Akbar, and A. S. Rachman, "Forecasting Produksi Energi Photovoltaic Menggunakan Algoritma Random Forest Classification," *Journal of Information System Research (JOSH)*, vol. 5, no. 4, pp. 1053–1062, Jul. 2024, doi: 10.47065/josh.v5i4.5514.
- [6] F. N. Farida, A. Faqih, and S. E. Permana, "Penerapan Model Prediksi Penjualan Pada Usaha Rumah Makan Menggunakan Algoritma Random Forest," *Jurnal Mahasiswa Teknik Informatika*, vol. 9, no. 4, pp. 5895–5901, Aug. 2025, doi: 10.36040/jati.v9i4.13912.
- [7] A. A. Wijanarko and H. Imaduddin, "Implementasi Metode Random Forest Untuk Memprediksi Jumlah Penjualan Gorden Berdasarkan Data Historis," *Journal of Information System Research*, vol. 7, no. 2, pp. 371–381, Jan. 2026, doi: 10.47065/josh.v7i2.9194.
- [8] G. Ramadhan, A. Faqih, and F. E. Permana, "Analisis Tren Penjualan Menu Seafood Dengan Algoritma Random Forest Untuk Meningkatkan Strategi Pemasaran," *Jurnal Mahasiswa Teknik Informatika*, vol. 9, no. 4, pp. 5942–5948, Aug. 2025, doi: 10.36040/jati.v9i4.13911.
- [9] B. C. Nusri, M. Yunus, N. Sirimorok, S. Alam, M. Zainal, and U. M. Parepare, "Implementasi Sistem Prediksi Pemesanan Tiket Online Untuk Optimalisasi Penjualan Menggunakan Random Forest," *JIKOM: Jurnal Informatika dan Komputer*, vol. 16, no. 1, pp. 55–63, Apr. 2026, doi: 10.46306/sm.v5i2.279.
- [10] D. Amiruddin, "Transformasi Data Transaksi Ke Deret Waktu Dan Evaluasi Model Peramalan Permintaan Pada Marketplace Plaza Banten," *Jurnal Innovation and Future Technology (IFTECH) P-ISSN*, vol. 8, no. 1, pp. 2656–2774, Feb. 2026, doi: 10.47080/iftech.v8i1.4550.
- [11] M. Erkamim, S. Suswadi, M. Z. Subarkah, and E. Widarti, "Komparasi Algoritma Random Forest dan XGBoosting dalam Klasifikasi Performa UMKM," *Jurnal Sistem Informasi Bisnis*, vol. 13, no. 2, pp. 127–134, Oct. 2023, doi: 10.21456/vol13iss2pp127-134.
- [12] A. N. Oktaviani et al., "Analisis Prediksi Radiasi Matahari Dengan Algoritma Machine Learning Dan Implementasi Bayesian Optimization Di Provinsi Dki Jakarta," *Jurnal Manajemen Informatika & Sistem Informasi (MISI)*, vol. 8, no. 1, Jan. 2025, doi: 10.36595/misi.v5i2.
- [13] G. A. Syafarina and Zaenuddin, "Prediksi Klik Iklan Online Menggunakan Regresi Logistik: Studi Empiris Tentang Pendekatan Berbasis Data," *Technologia: Jurnal Ilmiah*, vol. 16, no. 2, pp. 226–231, Apr. 2025, doi: 10.31602/tji.v16i2.15943.
- [14] J. Elektronika and D. Komputer, "Mengoptimalkan Proses Pembersihan Data dalam Analisis Big Data Menggunakan Pipeline Berbasis AI," *Jurnal Elektronika dan Komputer*, vol. 17, no. 2, pp. 657–666, Dec. 2024, doi: 10.51903/elkom.v17i2.2311.
- [15] A. Riansah, O. Nurdiawan, and R. Herdiana, "Penerapan Algoritma Random Forest dan Decision Tree untuk Meningkatkan Akurasi Klasifikasi Penjualan pada Toko Bangunan," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 9, no. 3, pp. 4242–4249, Jun. 2025, doi: <https://doi.org/10.36040/jati.v9i3.13622>.
- [16] B. Sandi and R. Muliono, "The Implementation Of Random Forest To Predict Sales A Case Study At Chatime Binjai Supermall," *JITE (Journal of Informatics and Telecommunication Engineering)*, vol. 8, no. 3, pp. 51–58, May 2025, doi: 10.31289/jite.v8i3Sp.14431.
- [17] A. Agustina, T. Tukino, B. Huda, and E. Novalia, "Prediksi Volume Penjualan Gadget Berdasarkan Promo dan Channel Penjualan Menggunakan Random Forest," *JUSIFOR: Jurnal Sistem Informasi dan Informatika*, vol. 4, no. 1, pp. 85–91, Jun. 2025, doi: 10.70609/jusifor.v4i1.6962.