

ANALISIS PREDIKSI PENENTUAN WAKTU DAN JUMLAH STOK PRODUK DALAM PENJUALAN MENGGUNAKAN ALGORITMA *XGBOOST REGRESSION*

Muhamad Insan Taufik
Fakultas Rekayasa Industri
Universitas Telkom
Bandung, Indonesia

insantaufik@student.telkomuniversity.ac.id

Irfan Darmawan
Fakultas Rekayasa Industri
Universitas Telkom
Bandung, Indonesia

irfandarmawan@telkomuniversity.ac.id

Oktariani Nurul Pratiwi
Fakultas Rekayasa Industri
Universitas Telkom
Bandung, Indonesia

onurulp@telkomuniversity.ac.id

Abstrak— Penelitian ini mengembangkan sistem *predictive analytics* untuk peramalan penjualan guna mendukung UMKM di Shopee dalam memperkirakan permintaan produk di masa mendatang. Sistem ini memungkinkan pelaku usaha untuk menentukan waktu dan jumlah pengisian ulang stok yang optimal, sehingga dapat mengatasi tantangan terkait kelebihan dan kekurangan stok yang berdampak pada biaya serta kepuasan pelanggan. Dengan memanfaatkan algoritma *XGBoost Regression* dalam metodologi *CRISP-DM* dan data penjualan historis, model dievaluasi menggunakan metrik MAE, RMSE, R^2 , dan MAPE, dan menunjukkan performa yang kuat ($R^2 = 0,901$, RMSE = 3,316, MAE = 0,873, dan MAPE < 10%). Hasilnya disajikan melalui grafik regresi yang membandingkan nilai aktual dan prediksi, garis *model-fit*, serta tabel prediksi yang menampilkan prakiraan penjualan bulanan untuk setiap produk selama periode 12 bulan. Pendekatan analisis prediktif ini terbukti dapat meningkatkan efisiensi manajemen persediaan dan mendukung pengambilan keputusan berbasis data bagi UMKM.

Kata Kunci— Analitik Prediktif, CRISP-DM, Pengisian Ulang Stok, Peramalan Penjualan, Regresi *XGBoost*, UMKM

I. PENDAHULUAN

RumahBayiTAZ Babyshop, sebuah UMKM di platform Shopee dengan pengalaman lebih dari delapan tahun dan memiliki 212 pengikut, berspesialisasi dalam kebutuhan bayi seperti popok, pakaian, empeng, pakaian renang, dan mainan edukatif. Beroperasi di pasar lokal dengan segmen khusus (*niche*), toko ini sepenuhnya mengandalkan saluran daring, namun menghadapi keterbatasan umum yang dialami UMKM, terutama dalam hal sumber daya yang terbatas untuk pengelolaan persediaan yang optimal.

Tantangan pengelolaan persediaan di RumahBayiTAZ muncul dalam bentuk *overstock* yakni kelebihan unit yang menumpuk di Gudang dan *understock*, yaitu kekosongan stok produk dengan permintaan tinggi seperti popok atau empeng saat periode promosi (Fig. 1). *Overstock* menyebabkan peningkatan biaya penyimpanan dan mengganggu arus kas, sementara *understock* berdampak pada hilangnya penjualan, ulasan negatif, serta merusak reputasi toko. Oleh karena itu, akurasi dalam waktu dan jumlah pengisian ulang stok menjadi sangat krusial untuk menyeimbangkan ketersediaan produk dengan efisiensi biaya. [1].

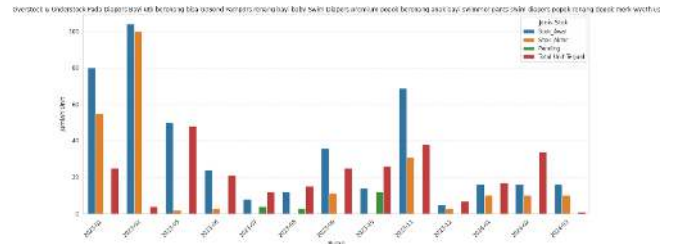


Fig. 1. Grafik Permasalahan Understock dan Overstock di RumahBayiTAZ Babyshop

Grafik ini menggambarkan distribusi bulanan stok dan penjualan untuk produk “*Diapers Bayi utk berenang bisa G0Send Pamp3rs renang bayi baby Swim Diapers premium popok berenang anak bayi swimmer pants swim diapers popok renang depok merk wyeth us*” di RumahBayiTAZ Babyshop dari Januari 2023 hingga Maret 2024. Sumbu X menunjukkan bulan, sementara sumbu Y menunjukkan jumlah unit. Terdapat empat kategori yang divisualisasikan: Stok Awal (biru), Stok Akhir (oranye), Pesanan Tertunda (hijau), dan Jumlah Terjual (merah).

Untuk mengatasi permasalahan tersebut, penelitian ini menerapkan analitik prediktif berbasis data dengan memanfaatkan model regresi *XGBoost* yang dikenal mampu menangkap interaksi fitur yang kompleks dan menghasilkan prediksi yang akurat. Dengan menganalisis data penjualan historis, pola musiman, dan waktu tunggu pemasok, algoritma ini memprediksi jadwal dan jumlah pengisian ulang stok yang optimal, sehingga dapat mengurangi risiko *overstock* dan *understock*. Tujuan utama dari penelitian ini adalah membangun kerangka prediktif yang mampu meningkatkan profitabilitas dan kepuasan pelanggan di RumahBayiTAZ Babyshop..

II. KAJIAN TEORI

A. *Overstock* dan *Understock*

Overstock terjadi ketika persediaan melebihi kebutuhan operasional akibat pemesanan berlebih atau kesalahan dalam peramalan, yang menyebabkan tingginya biaya penyimpanan serta risiko kedaluwarsa atau usangnya produk [2]. Model *Newsvendor* menggambarkan hal ini melalui *overage cost* (Co), yaitu kerugian per unit yang tidak terjual [3], [4]. Sebaliknya, *understock* mengacu pada ketidakcukupan persediaan untuk memenuhi permintaan, yang berujung pada kehilangan penjualan dan ketidakpuasan pelanggan. Kondisi

ini direpresentasikan oleh *underage cost* (C_u), yaitu margin keuntungan yang hilang akibat permintaan yang tidak terpenuhi [5].

B. Analitik Prediktif

Analitik prediktif memanfaatkan data historis dan pembelajaran mesin (*machine learning*) untuk

memperkirakan hasil di masa depan dengan memodelkan hubungan antara variabel input dan variabel target [6], [7], [8]. Pendekatan ini mendukung pengambilan keputusan dengan mengidentifikasi pola dalam kumpulan data berdimensi besar dan beragam, dengan mempertimbangkan faktor-faktor seperti kecepatan (*velocity*) dan kebenaran data (*veracity*) [9].

C. CRISP-DM

Model CRISP-DM membimbing proses pengembangan melalui tahapan pemahaman bisnis, persiapan data, pemodelan, dan evaluasi [10]. Model ini secara luas digunakan untuk menyusun tugas-tugas data mining dan selaras dengan alur metodologis dalam penelitian ini.

D. XGBoost dalam Regresi

XGBoost merupakan algoritma *gradient boosting* yang dioptimalkan dengan membangun serangkaian pohon keputusan (*decision tree*), di mana setiap pohon bertugas memperbaiki kesalahan dari pohon sebelumnya [11]. Algoritma ini unggul dalam menangani hubungan *nonlinier*, memiliki akurasi prediktif yang tinggi, bersifat skalabel untuk data berskala besar, dan dilengkapi dengan mekanisme regularisasi guna mengurangi risiko *overfitting*.

E. Extreme Gradient Boosting (Xgboost)

XGBoost (*Extreme Gradient Boosting*) merupakan kerangka kerja *boosting* berbasis pohon keputusan yang efisien, fleksibel, dan skalabel, dirancang untuk tugas klasifikasi dan regresi berkinerja tinggi. Algoritma ini merupakan implementasi yang dioptimalkan dari metode *gradient boosting*, yang dilengkapi dengan berbagai peningkatan di tingkat sistem untuk kecepatan dan akurasi yang lebih baik.

XGBoost menyempurnakan metode *boosting* tradisional dengan memberikan penalti secara eksplisit terhadap kompleksitas model dalam fungsi objektifnya. Alih-alih hanya meminimalkan *training loss*, *XGBoost* menambahkan komponen regularisasi yang mencegah model menghasilkan pohon yang terlalu besar atau bobot daun (*leaf weights*) yang ekstrem. Secara formal, fungsi objektifnya didefinisikan sebagai berikut:

$$L(\Phi) = \sum^n l(\hat{y}_i, y_i) + \sum^K \Omega(f_k) \quad (1)$$

Dengan istilah regularisasi $\Omega(f) = \gamma T + \frac{1}{2} \lambda |w|^2$

mana $l(\hat{y}_i, y_i)$ mengukur perbedaan antara nilai prediksi \hat{y}_i dan nilai aktual y_i , Istilah kedua $\Omega(f_k)$ meregulasi kompleksitas dari setiap pohon f_k . T_k adalah jumlah daun pada pohon ke- k , dan $w_{k,j}$ adalah bobot pada daun ke- j . Istilah γT_k bertindak sebagai semacam "biaya per daun", sementara istilah $\frac{1}{2} \lambda \sum w^2$ memberikan penalti terhadap

bobot daun yang terlalu besar, yang dapat menyebabkan *overfitting*.

Untuk mengoptimalkan $L(\Phi)$ secara efisien, *XGBoost* menggunakan pendekatan *stage-wise* dengan menambahkan satu pohon pada setiap iterasi. Pada setiap iterasi *boosting* ke- t , sebuah fungsi baru f_t dipelajari untuk meminimalkan fungsi objektif yang diperbarui. Karena meminimalkan *loss* secara langsung memerlukan komputasi yang tinggi, *XGBoost* mendekati fungsi objektif tersebut menggunakan pendekatan ekspansi Taylor orde dua (*second-order Taylor expansion*):

$$L^{(t)} \approx \sum_{i=1}^n [g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i)] + \Omega(f) \quad (2)$$

Dimana $L^{(t)}$ adalah fungsi kerugian (*loss*) pada iterasi ke- t , dan dalam persamaan tersebut, $g_i = \frac{\partial l(y_i, \hat{y}_i)}{\partial \hat{y}_i}$ merupakan

gradien yang menunjukkan arah pembaruan, sedangkan $h_i = \frac{\partial^2 l(y_i, \hat{y}_i)}{\partial \hat{y}_i^2}$ adalah Hessian yang menunjukkan ukuran langkah

(*step size*). Istilah kompleksitas tetap sesuai dengan definisi awal, yaitu $\Omega(f_t) = \gamma T_t + \frac{1}{2} \lambda \sum_{j=1}^{T_t} w_{t,j}^2$ tetap konsisten

dengan definisi awal. Dengan memanfaatkan turunan orde pertama (g_i) dan turunan orde kedua (h_i), *XGBoost* dapat melakukan pembaruan model secara lebih akurat dan stabil dibandingkan model yang hanya mengandalkan gradien saja. Keseimbangan antara akurasi prediktif, regularisasi, dan efisiensi optimasi ini menjadikan *XGBoost* sangat efektif, terutama ketika diterapkan pada dataset dunia nyata yang kompleks [11].

F. Data Preprocessing dan Feature Engineering

Pra-pemrosesan (*preprocessing*) mencakup proses pembersihan, pengkodean, transformasi, dan integrasi data dari berbagai sumber untuk memastikan data siap digunakan oleh model [12], [13], [14]. Generalisasi menyederhanakan atribut-atribut yang terlalu rinci, sementara integrasi menyelaraskan format antar dataset. Penanganan data hilang dilakukan melalui *imputation* (pengisian nilai) atau *deletion* (penghapusan) untuk menghindari bias. Rekayasa fitur (*feature engineering*) menciptakan variabel-variabel informatif seperti *indeks musiman* atau metrik penjualan terdahulu (*lagged sales*), yang sangat penting dalam prediksi *time-series* [15], [16].

G. Feature Selection

Seleksi fitur meningkatkan performa model dengan mempertahankan variabel-variabel yang memiliki daya prediktif tertinggi. Dalam penelitian ini digunakan metode korelasi peringkat Spearman, yaitu metode non-parametrik yang digunakan untuk mengidentifikasi hubungan monoton dan mengurangi redundansi antar fitur [17], [18]. Berbeda dengan korelasi Pearson yang mengasumsikan hubungan linear dan distribusi normal, Spearman bekerja berdasarkan peringkat nilai data, sehingga lebih tahan terhadap *outlier* dan

cocok untuk hubungan non-linear.

ISSN: 2355-9365 Di sini $L(\theta)$ merupakan total fungsi objektif yang akan diminimalkan. Istilah pertama adalah *empirical loss*, di

e-Proceeding of Engineering: Vol.13, No.1, Februari 2026 | Page 142 Metode seleksi fitur ini menggunakan rumus sebagai berikut :

$$\rho = 1 - \frac{\sum_i^2}{n(n^2-1)} \quad (3)$$

H. Pemodelan dan Evaluasi

Teknik *supervised learning* digunakan untuk tugas regresi, dengan performa model dievaluasi menggunakan metrik MAE, RMSE, MAPE, dan R^2 [13], [15]. Metrik-metrik ini mencerminkan baik besarnya kesalahan prediksi maupun kemampuan penjelasan model. Pendekatan evaluasi yang andal menggunakan cross-validation untuk mengurangi risiko *overfitting*.

MAE, RMSE, dan MAPE merupakan metrik utama untuk mengevaluasi akurasi prediktif, sedangkan R^2 mengukur proporsi variansi yang dapat dijelaskan oleh model [20], [21], [22], [23], [24]. MAPE sangat berguna karena skala persentasenya yang intuitif dan tidak bergantung pada satuan data [25], [26], [27].

R-squared (R^2) mengukur proporsi variansi dalam variabel target yang dapat dijelaskan oleh model. Nilainya berkisar antara 0 hingga 1, di mana nilai yang lebih tinggi menunjukkan kesesuaian model yang lebih baik. Secara matematis, R^2 didefinisikan sebagai berikut :

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

Root Mean Squared Error (RMSE) mengukur rata-rata besar kesalahan prediksi, dengan penalti yang lebih besar terhadap kesalahan yang besar karena adanya pemangkatan kuadrat. RMSE sensitif terhadap outlier dan digunakan untuk menilai seberapa jauh nilai prediksi menyimpang dari nilai aktual. Rumusnya adalah sebagai berikut :

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

Mean Absolute Error (MAE) menghitung rata-rata dari selisih absolut antara nilai prediksi dan nilai aktual. MAE memberikan interpretasi yang langsung dan mudah dipahami karena kesalahan dinyatakan dalam satuan yang sama dengan data. Rumusnya sebagai berikut :

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (6)$$

Mean Absolute Percentage Error (MAPE) menyatakan kesalahan prediksi dalam bentuk persentase terhadap nilai aktual, menjadikannya tidak bergantung pada skala data dan mudah dipahami secara intuitif [25], [26], [27]. Rumusnya adalah :

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (7)$$

III. METODE

A. Pengumpulan Data

Penelitian ini menggunakan lima dataset yang mencakup informasi transaksi, stok, dan hari libur dari RumahBayiTAZ Babyshop (*Shopee*) serta sumber resmi. Dataset tersebut meliputi catatan pesanan, *level* stok, performa penjualan, data variasi produk, dan kalender hari libur nasional yang

B. Data Processing

Data dibersihkan dengan menghapus duplikasi dan mengimputasi nilai yang hilang. Fitur temporal seperti bulan dan hari dalam minggu, variabel lag (misalnya penjualan 1/3/7

hari sebelumnya), serta indikator musiman dikonstruksi untuk menangkap pola waktu. Data kemudian diagregasi pada tingkat harian dan bulanan. Proses *feature engineering* mencakup pengkodean variabel kategorikal dan penskalaan nilai numerik. Setelah prapemrosesan selesai, model regresi dilatih menggunakan algoritma XGBoost pada dataset yang telah diproses.

C. Evaluasi

Dataset dibagi menjadi data latih dan data uji. Performa model dievaluasi menggunakan empat metrik regresi, yaitu: *R-squared* (R^2) (4), *Mean Absolute Error* (MAE) (6), *Root Mean Squared Error* (RMSE) (5), dan *Mean Absolute Percentage Error* (MAPE) (7), sebagaimana telah dibahas dalam Tinjauan Pustaka. Teknik visualisasi seperti plot prediksi vs aktual dan grafik residual digunakan untuk memverifikasi reliabilitas model.

D. Deployment

Proses deployment akan diwujudkan dalam bentuk aplikasi antarmuka berbasis Streamlit yang secara otomatis

melakukan pemrosesan kelima dataset (penjualan, status stok, stok toko, pesanan siap kirim, dan hari libur), memuat model XGBoost, menjalankan prediksi 12 bulan ke depan, serta menampilkan hasilnya dalam format tabel dan grafik interaktif yang dapat langsung diunduh sebagai file CSV.

IV. HASIL DAN PEMBAHASAN

A. Business Understanding

Penelitian ini bertujuan untuk membangun model prediktif guna memperkirakan waktu dan jumlah restock yang optimal di RumahBayiTAZ Babyshop, dalam rangka mengatasi risiko *overstock* dan *understock*. Permasalahan inventori sepanjang tahun 2023–2024 telah memengaruhi efisiensi operasional dan kepuasan pelanggan akibat prediksi penjualan yang tidak akurat, keterlambatan *restock*, serta

penumpukan stok produk dengan permintaan rendah. Dengan

memanfaatkan data historis, model prediktif dapat membantu mengidentifikasi pola permintaan dan tren musiman, sehingga mendukung pengambilan keputusan restocking yang lebih tepat.

A. Data Understanding

Lima dataset dari *Shopee Seller Centre* dan pemilik usaha

(periode Januari 2023 – Desember 2024) digunakan dalam penelitian ini, yang mencakup:

- Catatan transaksi (misalnya jumlah pesanan, harga).
- Level stok (awal dan akhir bulan).
- Metode performa toko (penjualan, kunjungan, retur).

mencakup periode tahun 2023–2024. Seluruh *file* telah distandarisasi dan digabungkan menggunakan *library pandas* dari *Python* untuk keperluan analisis lebih lanjut.

- Kalender hari libur (libur nasional dan cuti bersama).

Seluruh dataset tersebut telah disusun secara terstruktur dan digabungkan ke dalam satu file terpusat untuk keperluan analisis.

B. Data Processing

1) Pembersihan dan Generalisasi

Kolom-kolom yang redundan serta memiliki banyak nilai kosong dihapus. Nilai *null* pada nama variasi diberi label "Tidak ada variasi". Nama produk dan label variasi distandarisasi, misalnya dari "0-3 bln, Boy" menjadi "Usia 0-3 Bulan, Laki-laki".

TABLE I. HASIL PEMBERSIHAN DATA

Dataset	Perubahan	Hasil Pembersihan (Baris × Kolom)
Penjualan	Hapus kolom, hapus baris, Pengisian NA/ <i>null</i>	966 × 19
Pesanan Siap Dikirim	Tidak berubah	641 × 16
Status Stok Penjualan	Tidak berubah	357 × 7
Hari Libur	Tidak berubah	32 × 3
Stok Toko	Pengisian NA/ <i>null</i> pada kolom Nama Variasi	295 × 6

^a Ringkasan Perubahan Jumlah Baris × Kolom Setelah Pembersihan Data

TABLE II. GENERALIASI DATA PADA KOLOM NAMA PRODUK

^a Contoh dari perubahan di Kolom Nama Produk

TABLE III. GENERALIASI DATA PADA KOLOM NAMA VARIASI

^a Contoh dari perubahan di Kolom Nama Variasi

2) Integrasi Data

Dataset digabung berdasarkan variasi produk dan bulan, lalu dihitung agregat seperti penjualan total, kunjungan produk, dan perputaran stok. Ditambahkan pula indikator biner (*is_holiday_month*) untuk menandai bulan dengan hari libur.

3) Transformasi Data dan Feature Engineering

Variabel dikonversi ke format numerik atau di-*encoding* (misalnya *label encoding* untuk status stok). Lebih dari 40 fitur baru direkayasa, termasuk di antaranya:

- Fitur waktu: bulan, kuartal, indeks musiman
- Lag dan rata-rata bergerak.
- Pertumbuhan bulanan (MoM), perputaran stok, dan tingkat harga.

Fitur-fitur ini dirancang untuk menangkap pola permintaan jangka pendek maupun musiman.

4) Feature Selection

Korelasi *Spearman* digunakan untuk menilai fitur-fitur paling prediktif, termasuk variabel *lag*, rata-rata bergerak, laju pertumbuhan, dan indikator stok pada tingkat produk. Fitur-fitur teratas menunjukkan korelasi tinggi terhadap penjualan bulanan (misalnya, *Rolling_3M_Total_Jumlah*: 0,908).

Berikut adalah hasil analisis *feature importance* menggunakan rumus korelasi *Spearman* (3):

TABLE IV. HASIL KORELASI SPEARMAN

^a 9 Fitur Terbaik Berdasarkan Metode Korelasi Spearman

B. Pemodelan

Model regresi *XGBoost* dilatih terlebih dahulu menggunakan parameter default, kemudian disempurnakan melalui *GridSearchCV* dengan *TimeSeriesSplit*. Proses penyetulan ini berhasil meningkatkan performa model di semua metrik evaluasi:

TABLE V. PERBANDINGAN KINERJA MODEL *DEFAULT* DAN *TUNED*

^a Perbandingan model default dengan model tuned

Model yang telah disempurnakan digunakan untuk prediksi akhir karena memiliki akurasi yang lebih unggul.

C. Evaluasi

Model yang telah disempurnakan mencapai nilai R^2 sebesar 0,901, yang menunjukkan bahwa model mampu menjelaskan 90,1% variabilitas pada penjualan produk bulanan. Dengan nilai $RMSE = 3,316$ dan $MAE = 0,873$, model menunjukkan tingkat galat prediksi yang rendah, sedangkan $MAPE = 9,76\%$ mencerminkan ketepatan peramalan yang tinggi.

TABLE VI. INTERPRETASI METRIK

No	Metrik	Hasil	Interpretasi
1	R ²	0.902	Model ini mampu menjelaskan 90,2% variansi dari target.
2	RMSE	3.193	Rata-rata galat prediksi (RMSE) adalah sebesar 3,193.
3	MAE	0.907	Perbedaan absolut rata-rata antara nilai prediksi dan aktual (MAE) sebesar 0,907 menunjukkan tingkat akurasi yang baik.
4	MAPE	10,328 (10.3%)	Nilai Mean Absolute Percentage Error (MAPE) sebesar 10,328% dikategorikan sangat baik.

^a. Rincian dari nilai metrik model

Setelah model XGBoost menunjukkan performa baik ($R^2 = 0,902$; $RMSE = 3,193$; $MAPE = 10,328\%$; $MAE = 0,907$), model digunakan untuk memprediksi jumlah penjualan selama 12 bulan ke depan untuk tiap produk. Sebelum evaluasi, dijelaskan terlebih dahulu grafik hasil prediksi yang terdiri dari: garis biru (data aktual), garis oranye putus-putus (*model fit*), dan garis oranye solid (*forecast*). Grafik ini membantu menganalisis efektivitas model dalam menyelesaikan masalah *overstock* dan *understock* pada penelitian ini.

Fig. 2. Hasil Visualisasi Grafik Prediksi

Berdasarkan Fig. 2, Grafik *model fit* menunjukkan prediksi model sangat mendekati data aktual, terutama pada lonjakan penjualan Mei–Juni dan November. Prediksi 12 bulan ke depan juga konsisten menampilkan pola permintaan puncak dan rendah, yang mendukung perencanaan *restock* yang lebih akurat.

Dalam konteks manajemen inventaris, model ini membantu penetapan indikator penting seperti *lead time*, *reorder point*, *safety stock*, dan *service level*. Model memfasilitasi perhitungan stok optimal berdasarkan prediksi permintaan dan variabel penting seperti bulan, promosi, dan hari libur. Hasilnya, tim operasional dapat mengambil keputusan berbasis data untuk menghindari risiko *overstock* maupun *understock*, sekaligus meningkatkan efisiensi rantai pasok.

D. Deployment

Dalam kasus prediksi ini, proses *deployment* dilakukan dengan memanfaatkan aplikasi *web* yang dibangun menggunakan *Streamlit*. Model prediksi penjualan yang telah dilatih sebelumnya disimpan dalam format *.pkl* dan dimuat ulang menggunakan modul *joblib*. Aplikasi memungkinkan pengguna untuk mengunggah lima *dataset* pendukung, yang kemudian diproses dan ditransformasi secara otomatis sesuai dengan *pipeline* yang telah dirancang. Setelah pengguna menekan tombol "*Proses Data & Generate Prediksi*", hasil prediksi selama 12 bulan ke depan ditampilkan secara dinamis melalui grafik interaktif, tabel, dan opsi *ekspor* dalam format *CSV*. *Antarmuka* dirancang responsif dan terstruktur menggunakan fitur *st.columns()* dan *tab navigasi*, sehingga memudahkan pengguna dalam mengeksplorasi informasi prediktif dari berbagai produk secara efisien.

Fig. 3. Desain Antarmuka Fitur Unggah Data

Fig. 4. Desain Antarmuka Fitur Hasil Prediksi (Tabel)

Fig. 5. Desain Antarmuka Fitur Hasil Prediksi (Grafik)

Berdasarkan tampilan pada Fig. 3. hingga Fig. 5., sistem prediksi yang dibangun menunjukkan proses dan hasil yang sangat sistematis dan responsif. Fig. 3. memperlihatkan

antarmuka unggah data yang mendukung lima *dataset* utama. Indikator status sistem seperti "Backend Aktif", "Model Tersedia", dan "Data Siap" menunjukkan bahwa sistem dalam kondisi siap menjalankan prediksi setelah semua file berhasil diunggah.

Selanjutnya, pada *Fig. 4.*, sistem memungkinkan pengguna memilih produk tertentu dan melakukan konfigurasi prediksi berdasarkan jumlah bulan ke depan. Hasil prediksi ditampilkan dalam bentuk tabel yang menyajikan jumlah penjualan bulanan selama 12 bulan ke depan. Indikator keberhasilan prediksi juga muncul secara visual, memberikan umpan balik positif yang informatif.

Fig. 5. menampilkan dua grafik utama: perbandingan antara data historis dan hasil prediksi (*line chart*), serta rincian prediksi untuk produk tertentu. Grafik ini menunjukkan bahwa pola prediksi mengikuti tren data sebelumnya secara wajar. Nilai total prediksi (171 unit) dan rata-rata penjualan bulanan (14 unit) disajikan dengan jelas. *Visualisasi* ini membuktikan bahwa sistem mampu menghasilkan *output* yang tidak hanya akurat tetapi juga mudah diinterpretasikan oleh pengguna secara praktis.

V. KESIMPULAN

Berdasarkan keseluruhan tahapan penelitian, dapat disimpulkan bahwa pemodelan penjualan menggunakan *XGBoost* yang dilengkapi fitur temporal, *lag*, *rolling*, dan hari libur mampu menghasilkan prediksi yang akurat dengan nilai R^2 sebesar 0,902 dan MAE sebesar 0,907. Model ini berhasil menangkap tren musiman dan pola jangka panjang penjualan untuk 12 bulan ke depan.

Prediksi yang dihasilkan memungkinkan perencanaan waktu dan jumlah *restock* secara optimal, dengan mengidentifikasi periode puncak dan penurunan permintaan. Hal ini membantu meminimalkan risiko *overstock* dan *understock* melalui proyeksi kuantitatif yang andal.

Integrasi model ke dalam aplikasi *Streamlit* dengan visualisasi interaktif memperkuat penerapan hasil prediksi dalam konteks operasional. Fitur grafik dan tabel prediksi memberikan kemudahan interpretasi bagi pengguna dalam pengambilan keputusan berbasis data.

REFERENSI

- [1] Max Muller, "Essentials of Inventory Management," Jan. 2019, Accessed: Dec. 31, 2024. [Online]. Available: https://www.google.co.id/books/edition/Essentials_of_Inventory_Management/R_JWDwAAQBAJ?hl=id&gbpv=1&dq=Essentials+of+Inventory+Management&pg=PR13&printsec=frontcover
- [2] D. R. Hansen and M. M. Mowen, *Managerial Accounting*, 8th ed. Mason, OH: Thomson South-Western, 2007.
- [3] F. Y. Edgeworth, "The mathematical theory of banking," *Journal of the Royal Statistical Society*, vol. 51, no. 1, pp. 113–127, 1888, doi: 10.2307/2979336.
- [4] K. J. Arrow, T. Harris, and J. Marschak, "Optimal inventory policy," *Econometrica*, vol. 19, no. 3, pp. 250–272, 1951, doi: 10.2307/1906813.
- [5] S. Chopra and P. Meindl, *Supply Chain Management: Strategy, Planning, and Operation*, 6th ed. Boston: Pearson Education, 2016.
- [6] F. Drasgow, "Predictive HR Analytics: Mastering the HR Metric (2nd Ed.)." Martin R. Edwards and Kirsten Edwards, New York: Kogan Page Ltd., 2019, 515 pages, \$136 hardcover, \$42.31 paperback.,” *Pers Psychol*, vol. 72, no. 4, pp. 631–633, Dec. 2019, doi: 10.1111/peps.12361.
- [7] J. F. Hair, M. Page, and N. Brunsveld, *Essentials of Business Research Methods*. Routledge, 2019. doi: 10.4324/9780429203374.
- [8] S. Geisser, *Predictive Inference: An Introduction*. Boston, MA: Springer US, 1993. doi: 10.1007/978-1-4899-4467-2.
- [9] J. Rajni and D. B. Malaya, "Predictive Analytics in a Higher Education Context," *IT Prof*, vol. 17, no. 4, pp. 24–33, Jul. 2015, doi: 10.1109/MITP.2015.68.
- [10] R. Wirth and J. Hipp, "CRISP-DM: towards a standard process model for data mining. Proceedings of the Fourth International Conference on the Practical Application of Knowledge Discovery and Data Mining, 29-39," *Proceedings of the Fourth International Conference on the Practical Application of Knowledge Discovery and Data Mining*, no. 24959, pp. 29–39, 2000, [Online]. Available: https://www.researchgate.net/publication/239585378_CRISP-DM_Towards_a_standard_process_model_for_data_mining
- [11] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, vol. 13-17-Aug, pp. 785–794, 2016, doi: 10.1145/2939672.2939785.
- [12] P. Chapman, "CRISP-DM 1.0: Step-by-step data mining guide," 2000. [Online]. Available: <https://api.semanticscholar.org/CorpusID:59777418>
- [13] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed. Morgan Kaufmann, 2012.
- [14] R. J. A. Little and D. B. Rubin, *Statistical Analysis with Missing Data*, 2nd ed. Wiley, 2002.
- [15] A. Zheng and A. Casari, *Feature Engineering for Machine Learning: Principles and Techniques for Data Scientists*. Sebastopol, CA: O'Reilly Media, Inc., 2018.
- [16] F. Chollet, *Deep Learning with Python*. Shelter Island, NY: Manning Publications, 2018.
- [17] L. Breiman, "Random forests," *Mach Learn*, vol. 45, no. 1, pp. 5–32, 2001.
- [18] S. García, J. Luengo, and F. Herrera, *Data Preprocessing in Data Mining*. Springer, 2015.
- [19] A. Paleyes, R.-G. Urma, and N. D. Lawrence, "Challenges in Deploying Machine Learning: A Survey of Case Studies," *arXiv preprint arXiv:2011.09926*, 2020, [Online]. Available: <https://arxiv.org/abs/2011.09926>
- [20] A. Botchkarev, "Performance Metrics (Error Measures) in Machine Learning Regression, Forecasting and Prognostics: Properties and Typology," *arXiv preprint arXiv:1809.03006*, 2018, [Online]. Available: <https://arxiv.org/abs/1809.03006>
- [21] K. Pearson, "On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 50, no. 302, pp. 157–175, 1900.
- [22] A.-M. Legendre, *Nouvelles méthodes pour la détermination des orbites des comètes*. Paris: Courcier, 1805.
- [23] C. F. Gauss, *Theoria motus corporum coelestium in sectionibus conicis solem ambientium*. Perthes et Besser, 1809.

- [24] S. M. Stigler, "Gauss and the Invention of Least Squares," *The Annals of Statistics*, vol. 9, no. 3, pp. 465–474, 1981.
- [25] R. J. Boscovich, "De litteraria expeditione per pontificiam ditionem ad dimetiendos duos meridiani gradus et corrigendam mappam geographicam," *Rome*, 1757.
- [26] B. L. Bowerman, R. T. O'Connell, and A. B. Koehler, *Forecasting, Time Series, and Regression: An Applied Approach*. Belmont, CA: Thomson Brooks/Cole, 2004.
- [27] S. Makridakis *et al.*, "The accuracy of extrapolation (time series) methods: Results of a forecasting competition," *J Forecast*, vol. 1, no. 2, pp. 111–153, 1982.