

## MODEL PREDIKSI OMZET PENJUALAN MENGGUNAKAN RANDOM FOREST REGRESSION PADA UMKM TOKO KEMBAR

Umi Fidya Hamidah, Aprilisa Arum Sari, Ratna Puspita Indah

Teknik Informatika, Universitas Duta Bangsa Surakarta  
Jl. Bhayangkara No. 55 kota Surakarta, Jawa Tengah, Indonesia  
220103191@mhs.udb.ac.id

### ABSTRAK

UMKM retail di Indonesia, termasuk Toko Kembar di Sukoharjo, masih mengandalkan pencatatan transaksi secara manual sehingga data penjualan belum dimanfaatkan secara optimal. Permasalahan utama yang dihadapi adalah ketidakmampuan memprediksi fluktuasi omzet penjualan, yang berdampak pada pengelolaan stok dan perencanaan keuangan yang reaktif. Penelitian ini bertujuan membangun model prediksi omzet penjualan berbasis *Random Forest Regression* untuk mendukung pengelolaan stok yang lebih terencana dan berbasis data. Metode yang digunakan adalah kerangka *CRISP-DM* dengan pendekatan panel data per produk per bulan, menggunakan rekayasa fitur berupa lag omzet 1–3 bulan, rata-rata bergerak 3 bulan, harga satuan, dan jumlah transaksi. Dataset mencakup 23.466 transaksi dari 8.720 produk selama periode September 2025–Maret 2026. Evaluasi model menggunakan *MAE*, *RMSE*,  $R^2$ , dan *MAPE*. Hasil evaluasi model menunjukkan nilai *MAE* sebesar Rp 32.579, *RMSE* sebesar 277.705,  $R^2$  sebesar 0,6702 (67,02%), dan *MAPE* sebesar 8,45%. Prediksi omzet April 2026 sebesar Rp 661.396.513 dan Mei 2026 sebesar Rp 651.661.400 dengan rekomendasi penyesuaian stok pada beberapa produk.

**Kata kunci :** *Random Forest Regression, Prediksi Omzet, UMKM, Machine Learning, CRISP-DM, Manajemen Stok*

### 1. PENDAHULUAN

UMKM (Usaha Mikro, Kecil, dan Menengah) retail merupakan tulang punggung perekonomian Indonesia. Berdasarkan data Kementerian Koperasi dan UKM, UMKM menyumbang lebih dari 60% Produk Domestik Bruto (PDB) nasional dan menyerap sebagian besar tenaga kerja. Di tengah kontribusinya yang besar, mayoritas UMKM retail masih mengandalkan pencatatan transaksi secara manual, sehingga data penjualan historis tidak terkelola dengan baik. Kondisi ini menyebabkan laporan keuangan yang rawan kesalahan serta sulitnya pelacakan kinerja usaha secara akurat [1]. Studi menunjukkan bahwa UMKM tidak dapat memprediksi pendapatan dan merencanakan strategi bisnis dengan baik karena kekurangan data penjualan historis[2].

Toko Kembar, sebuah usaha mikro kecil dan menengah (UMKM) juga telah menghadapi tantangan ini. Meskipun toko ini memiliki data historis penjualan sebelumnya, mereka belum menggunakannya sepenuhnya untuk memprediksi penjualan untuk membuat keputusan tentang pengelolaan stok di masa mendatang. Pemilik usaha kemudian menghadapi kesulitan dalam menentukan rencana pembelian barang, pengelolaan modal, dan perencanaan keuangan[3]. Pengambilan keputusan yang tidak berbasis data dan reaktif disebabkan oleh fluktuasi penjualan yang tidak terprediksi[4].

Prediksi penjualan yang dibuat dengan metode *machine learning*, khususnya *Random Forest*, telah dipelajari oleh banyak penelitian ini telah terbukti lebih akurat dan stabil daripada metode regresi konvensional. Namun, penelitian tersebut belum banyak diterapkan pada UMKM retail skala kecil

secara khusus dan biasanya berfokus pada prediksi jumlah penjualan daripada prediksi omzet sebagai indikator kinerja keuangan[5].

Berdasarkan celah tersebut, penelitian ini menawarkan metode analisis yang menggunakan algoritma *Random Forest regression* sebagai dasar untuk membuat keputusan tentang pengelolaan stok penjualan berdasarkan data historis Toko Kembar. Diharapkan metode ini dapat menghasilkan informasi prediktif dari data penjualan, yang akan membantu pengambilan keputusan keuangan[6]. Penelitian ini penting karena kebutuhan UMKM terhadap pengelolaan bisnis berbasis data semakin meningkat di era digital. Sebagai rencana penyelesaian masalah, penelitian ini mengusulkan penerapan algoritma *Random Forest Regression* dalam kerangka *CRISP-DM* untuk membangun model prediksi omzet berbasis data historis penjualan Toko Kembar periode September 2025–Maret 2026. Model ini dirancang untuk menghasilkan prediksi omzet 2 bulan ke depan serta rekomendasi stok per produk secara otomatis, yang dapat diakses pemilik toko melalui antarmuka aplikasi berbasis web tanpa keahlian pemrograman. Dengan demikian, diharapkan Toko Kembar dapat beralih dari pengelolaan stok yang reaktif menuju pengelolaan yang proaktif dan berbasis data, sehingga dapat meningkatkan efisiensi modal dan mengurangi risiko kerugian akibat kekurangan maupun kelebihan stok.

### 2. TINJAUAN PUSTAKA

#### 2.1. Penelitian Terdahulu

Penelitian yang telah dilakukan sebelumnya membandingkan dua algoritma pembelajaran mesin,

*Random Forest* dan *XGBoost*, untuk membuat model prediksi penjualan untuk *Coffee Shop OI*. Tujuan penelitian ini adalah untuk membantu mengelola stok bahan baku, yang selama ini dilakukan secara manual dan mengakibatkan pembelian yang tidak terencana. Dalam penelitian ini, data transaksi selama enam bulan digunakan untuk proses pelatihan dan pengujian model dengan evaluasi berdasarkan akurasi, *RMSE*, dan *MAPE*. Hasilnya menunjukkan bahwa dengan rasio data latih 80:20, *Random Forest* memberikan performa terbaik dengan akurasi 72% dan lebih stabil daripada *XGBoost*. Oleh karena itu, *Random Forest* dapat disarankan untuk pengelolaan bahan baku untuk mengurangi risiko kekurangan dan kelebihan persediaan[7].

Untuk mendukung strategi bisnis dan perencanaan investasi dalam industri pertambangan, penelitian sebelumnya menjelaskan tentang sebuah usaha mikro kecil dan menengah (UMKM) juga menghadapi tantangan ini. Walaupun usaha ini memiliki rekam jejak penjualan sebelumnya, Mempelajari seberapa tepat *Naive Bayes* dan *Random Forest* dalam memprediksi penjualan emas. Temuan dari penelitian ini mengindikasikan bahwa *Naive Bayes* cukup sukses dalam mengenali barang yang menarik minat pembeli. Sementara itu, *Random Forest*, karena kemampuan kelompoknya untuk menangani data kompleks dan mengurangi *overfitting*, lebih baik dalam memprediksi perilaku repeat order pada pelanggan. Meskipun data pengujian tidak selalu akurat, *Random Forest* masih mampu membuat prediksi yang kuat, yang menjadikannya sebagai opsi yang sangat tepat untuk model estimasi penjualan yang didasarkan pada data [8].

Dengan menggunakan algoritma *Random Forest*, penelitian sebelumnya juga mengembangkan model untuk memprediksi volume penjualan perangkat dengan mempertimbangkan promosi dan channel penjualan. Data yang digunakan mencakup penjualan elektronik di toko offline dan online, dan kemudian dikategorikan menjadi tiga kategori volume penjualan: rendah, sedang, dan tinggi. Hasil evaluasi menunjukkan bahwa model *Random Forest* memiliki akurasi sebesar 49,5% dan paling akurat dalam kategori penjualan "sedang". Hasil ini menunjukkan bahwa pembelajaran mesin dapat membantu strategi pemasaran berbasis data, meskipun akurasi masih perlu ditingkatkan [9].

## 2.2. Prediksi Omzet Penjualan

Dengan menggunakan data historis transaksi, prediksi omzet penjualan adalah proses memperkirakan pendapatan penjualan pada periode mendatang. Prediksi omzet penting untuk membantu pelaku usaha mengelola modal, stok, dan strategi bisnis. Dengan menggunakan prediksi omzet, pemilik usaha dapat membuat keputusan bisnis yang lebih tepat.

## 2.3. Machine Learning

*Machine learning* sangat populer di bidang bisnis, kesehatan, pendidikan, dan penjualan. Ini adalah cabang kecerdasan buatan yang memungkinkan sistem komputer tanpa diprogram secara eksplisit digunakan untuk mempelajari pola data untuk menghasilkan prediksi atau keputusan [10].

## 2.4. Random Forest Regression

Algoritma *ensemble learning* dikenal sebagai *Random Forest Regression* yang memanfaatkan output dari berbagai *decision tree*, algoritma ini mampu meningkatkan ketepatan prediksi serta mengurangi masalah *overfitting* [9].

## 2.5. CRISP-DM

*CRISP-DM* (*Cross Industry Standard Process for Data Mining*) adalah metode pengembangan data mining yang terdiri dari enam tahapan yaitu *business understanding*, *data understanding*, *data preparation*, *modeling*, *evaluation*, dan *deployment* [11].

## 3. METODE PENELITIAN

Metode yang digunakan untuk merancang model dalam studi ini adalah model *CRISP-DM* (*Cross Industry Standard Process for Data Mining*), yang meliputi enam tahapan yang saling berhubungan yang mencakup pemahaman bisnis dan implementasi model [11]. Seluruh prosedur analisis dilakukan dengan memanfaatkan bahasa pemrograman Python di platform Google Colab dengan *library scikit-learn*, *pandas*, *numpy*, dan *plotly*.

### 3.1. Business Understanding

Tujuan dari tahap ini adalah untuk memperoleh wawasan mengenai masalah bisnis yang sedang di Toko Kembar, yaitu fluktuasi penjualan yang tidak dapat diprediksi yang menyebabkan perencanaan keuangan dan pengelolaan stok tetap reaktif. Tujuan dari studi ini adalah menciptakan sebuah model untuk meramalkan penjualan yang berlandaskan pada informasi masa lalu dan dimanfaatkan sebagai acuan dalam membuat keputusan yang lebih akurat dan terukur terkait pengelolaan persediaan.

### 3.2. Data Understanding

Data yang digunakan berasal dari aplikasi sistem kasir toko Kembar selama periode 7 bulan, dari September 2025 hingga Maret 2026. Dataset mengandung 23.466 catatan transaksi yang memiliki variabel sebagai berikut: nama barang, kode barang, harga satuan, jumlah barang terjual (*quantity*), jumlah transaksi (*trx*), dan total omzet (jumlah). Dataset mengandung 8.720 produk unik. Pada titik ini, eksplorasi awal dilakukan untuk mendapatkan pemahaman tentang distribusi data, pola penjualan per periode, dan produk mana yang memiliki frekuensi transaksi tertinggi. Tercatat total omzet selama 7 bulan sebesar Rp 2.157.265.009.

### 3.3. Data Preparation

Tahap persiapan data meliputi: (1) pembersihan data (*data cleaning*) dengan menghapus nilai kosong dan memperbaiki format angka yang tidak konsisten, (2) normalisasi kolom numerik (*quantity*, harga, jumlah, *trx*), (3) agregasi data menjadi format panel per produk per bulan menggunakan *groupby*. Selanjutnya dilakukan rekayasa fitur (*feature engineering*) dengan membangun variabel-variabel: lag omzet bulan ke-1 (*omzet\_lag1*), lag omzet bulan ke-2 (*omzet\_lag2*), lag omzet bulan ke-3 (*omzet\_lag3*), rata-rata bergerak 3 bulan (*rata\_rata\_3bulan*), harga satuan rata-rata, *quantity* rata-rata, jumlah transaksi rata-rata, dan kode bulan numerik. Produk dengan data kurang dari 2 bulan dikeluarkan dari pemodelan karena tidak cukup untuk membangun fitur lag. Total data yang digunakan untuk pelatihan sebanyak 19.792 baris.

### 3.4. Modeling

Algoritma yang diterapkan adalah Random Forest Regression, yaitu pendekatan *ensemble learning* yang menciptakan beberapa *decision tree* secara paralel dan mengambil rata-rata prediksinya untuk memberikan hasil yang lebih konsisten dan tepat. Pembagian data menggunakan pendekatan *time-based split* bulan terakhir (Maret 2026) dijadikan data uji, sedangkan bulan September 2025 hingga Februari 2026 digunakan sebagai data latih. Optimasi hyperparameter dilakukan menggunakan *RandomizedSearchCV* dengan 10 iterasi dan 3-fold cross validation. Parameter yang dioptimasi meliputi: *n\_estimators* (100, 150, 200, 300), *max\_depth* (10, 15, 20, None), *min\_samples\_split* (2, 5, 10), dan *min\_samples\_leaf* (1, 2). Prediksi ke depan dilakukan secara iteratif untuk 2 bulan berikutnya (April dan Mei 2026) dengan memanfaatkan nilai prediksi bulan sebelumnya sebagai fitur lag pada prediksi bulan berikutnya.

### 3.5. Evaluation

Evaluasi model dilakukan menggunakan empat metrik: (1) *MAE (Mean Absolute Error)*, yaitu rata-rata selisih absolut antara nilai aktual dan prediksi per produk, (2) *RMSE (Root Mean Squared Error)*, yaitu akar dari rata-rata kuadrat selisih prediksi yang lebih sensitif terhadap *outlier*, (3) *R<sup>2</sup> Score (koefisien determinasi)*, yang mengukur proporsi variansi omzet yang dapat dijelaskan oleh model, dan (4) *MAPE (Mean Absolute Percentage Error)*, yang mengukur persentase kesalahan prediksi rata-rata per produk. Selain evaluasi pada level produk, dilakukan juga evaluasi total omzet dengan membandingkan selisih antara total omzet aktual dan total omzet prediksi pada bulan uji.

### 3.6. Deployment

Mengimplementasikan model prediksi yang telah dibangun ke dalam aplikasi berbasis Streamlit menggunakan Python melakukan tahap deployment.

Aplikasi ini memungkinkan pengguna mengunggah data penjualan, menjalankan model, dan memperoleh hasil prediksi omzet dan rekomendasi stok secara interaktif melalui antarmuka web tanpa harus memiliki keahlian pemrograman. Tabel rekomendasi stok dengan kategori aksi (tambah, pertahankan, atau kurangi) dan visualisasi omzet historis dan prediksi per produk adalah hasil yang dihasilkan. Mereka juga menghasilkan estimasi modal yang dibutuhkan untuk periode prediksi berikutnya.

## 4. HASIL DAN PEMBAHASAN

Hasil evaluasi model *Random Forest Regression* pada data uji (Maret 2026) disajikan secara rinci pada sub-bagian berikut. Model dibangun menggunakan pendekatan *ensemble learning* yang menggabungkan prediksi dari sejumlah *decision tree* yang dilatih secara paralel. Prinsip utama *Random Forest* adalah metode *bagging (Bootstrap Aggregating)*, di mana setiap pohon dilatih pada subset data yang diambil secara acak dengan pengembalian (*bootstrap sample*).

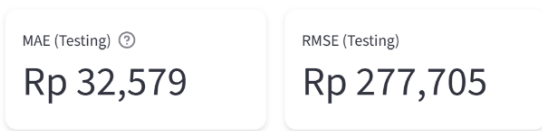
Pada setiap simpul pembelahan pohon, algoritma hanya mempertimbangkan subset fitur yang dipilih secara acak (*random feature selection*), sehingga setiap pohon menjadi lebih beragam dan korelasi antar pohon berkurang. Hasil prediksi akhir diperoleh dari rata-rata output seluruh pohon, yang terbukti mengurangi *variance* model dan meningkatkan generalisasi dibandingkan dengan pohon keputusan tunggal.

Pada penelitian ini, proses *hyperparameter tuning* dilakukan menggunakan *RandomizedSearchCV* dengan 10 iterasi dan 3-fold cross-validation. Parameter optimal yang diperoleh antara lain: *n\_estimators* = 200 (jumlah pohon), *max\_depth* = 15 (kedalaman maksimum pohon), *min\_samples\_split* = 5, dan *min\_samples\_leaf* = 2. Hasil evaluasi model pada data uji menunjukkan nilai MAE sebesar Rp 32.579 per produk, RMSE sebesar Rp 277.705, dan R<sup>2</sup> sebesar 0,6702 (67,02%). Nilai MAPE per produk mencapai 8,45%, sedangkan selisih total omzet antara aktual dan prediksi sebesar 18,71%. Total omzet aktual pada bulan uji adalah Rp 451.617.198, sementara prediksi model menghasilkan Rp 367.108.793. Nilai R<sup>2</sup> sebesar 67,02% menunjukkan model mampu menjelaskan 67% variasi omzet antar produk. Nilai MAE sebesar Rp 32.579 per produk mengindikasikan bahwa secara rata-rata, prediksi model meleset sekitar Rp 32.579 dari omzet aktual per produk per bulan.

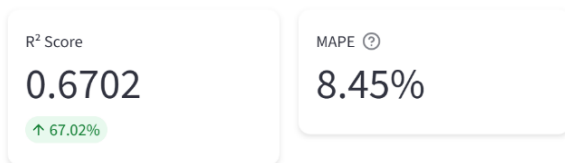
Ini merupakan tingkat kesalahan yang dapat diterima mengingat rata-rata omzet per produk per bulan pada dataset sangat bervariasi, dari produk bernilai puluhan ribu hingga jutaan rupiah. Nilai RMSE sebesar Rp 277.705 yang jauh lebih besar dari MAE mengindikasikan adanya sejumlah produk dengan kesalahan prediksi yang cukup besar (*outlier*), umumnya pada produk dengan omzet tinggi dan pola penjualan yang tidak konsisten. Nilai MAPE 8,45% dapat dikategorikan sebagai akurasi yang baik (di

bawah 10%) menurut standar evaluasi model prediksi. Selisih total omzet sebesar 18,71% dipengaruhi oleh produk-produk dengan volume transaksi tinggi yang sulit diprediksi karena keterbatasan data historis yang hanya tersedia 7 bulan.

### Metrik Evaluasi Model (Level Produk)

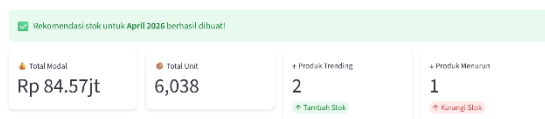


Gambar 1. Hasil evaluasi model



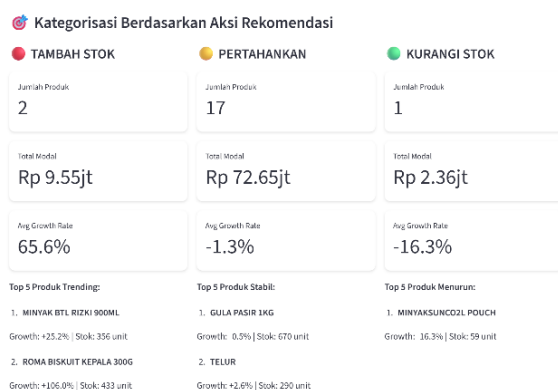
Gambar 2. Hasil evaluasi model

Performa ini cukup baik mengingat tingginya keragaman produk (8.720 item) dan keterbatasan periode historis 7 bulan. Total omzet historis yang tercatat adalah Rp 2.157.265.009. Prediksi omzet 2 bulan ke depan menunjukkan April 2026 sebesar Rp 661.396.513 dan Mei 2026 sebesar Rp 651.661.400, dengan total prediksi 2 bulan sebesar Rp 1.313.057.913.



Gambar 3. Hasil prediksi rekomendasi stok

Berdasarkan hasil prediksi tersebut, rekomendasi stok untuk periode April 2026 menunjukkan total modal yang dibutuhkan sebesar Rp 84.565.113, dengan 2 produk disarankan penambahan stok, 17 produk dipertahankan, dan 1 produk perlu pengurangan stok.



Gambar 4. Hasil prediksi stok produk

Produk dengan omzet tertinggi meliputi Gula Pasir 1kg (rekomendasi 670 unit, modal Rp 10.697.220), Telur (290 unit, modal Rp 7.602.350), dan Minyak BTL Rizki 900ml (356 unit, modal Rp 5.874.000) yang diprediksi naik 25,2% dari rata-rata historis sehingga masuk kategori tambah stok. Hasil ini menunjukkan model *Random Forest Regression* mampu menghasilkan informasi prediktif yang actionable untuk mendukung pengambilan keputusan pengelolaan stok pada UMKM Toko Kembar.

## 5. KESIMPULAN DAN SARAN

Dengan data historis dari September 2025 hingga Maret 2026, penelitian ini berhasil membangun model prediksi omzet penjualan menggunakan algoritma *Random Forest Regression*. Model menunjukkan kemampuan prediksi yang memadai untuk skala UMKM retail dengan ribuan produk dan periode historis terbatas; hasilnya adalah MAE sebesar 32.579, RMSE sebesar 277.705, R² sebesar 0,6702 (67,02%), dan MAPE sebesar 8,45% untuk setiap produk. Rekomendasi stok berbasis data yang dapat digunakan oleh pemilik toko untuk membuat keputusan keuangan yang lebih terencana disertai dengan perkiraan penjualan untuk April 2026 sebesar Rp 661.396.513 dan Mei 2026 sebesar Rp 651.661.400. Untuk menangkap pola musiman, penelitian lebih lanjut disarankan untuk memperpanjang periode historis setidaknya 12 bulan; menambahkan fitur dari luar, seperti hari libur dan momen pasar; dan mengembangkan antarmuka berbasis web yang memungkinkan pemilik UMKM yang tidak memiliki keahlian teknis untuk menggunakan model secara langsung.

## DAFTAR PUSTAKA

- [1] M. Syahrul Efendi *et al.*, “RESOLUSI: Rekayasa Teknik Informatika dan Informasi Penerapan Algoritma Random Forest Untuk Prediksi Penjualan Dan Sistem Persediaan Produk,” *Media Online*, vol. 5, no. 1, p. 20, 2024, doi: 10.30865/resolusi.v5i1.2149.
- [2] A. Sari, M. Arifin, and E. Darmanto, “Prediksi Kebutuhan Stok Barang Menggunakan Algoritma Random Forest untuk Meningkatkan Efisiensi Penjualan,” *Journal Of Information Systems And Informatics Engineering*, vol. 9, no. 2, pp. 339–351, 2025.
- [3] M. Adi Kurniawan, G. Zaidan Syauqi, M. Safriyanti, F. Ullul Azmie, and A. Setiawan, “JSI (Jurnal Sistem Informasi) Universitas Dirgantara Marsekal Suryadarma Prediksi Pendapatan Penjualan di Indomaret Menggunakan Algoritma Random Forest Regression,” *JSI (Jurnal Sistem Informasi) Universitas Dirgantara Marsekal Suryadarma*, 2025, [Online]. Available: <https://www.kaggle.com/datasets/firmanhermawan/data-sales-simulasi-indomaret>
- [4] A. Zulfia, T. N. Ilfa, Z. Damia, T. S. A. Sukiman, and A. Karima, “AI Decision Support for

- Demand Forecasting and Retail Stock Using Random Forest,” *Brilliance: Research of Artificial Intelligence*, vol. 5, no. 2, pp. 800–805, Aug. 2025, doi: 10.47709/brilliance.v5i2.5901.
- [5] B. Sandil and R. Muliono2, “JITE (Journal of Informatics and Telecommunication Engineering) The Implementation Of Random Forest To Predict Sales A Case Study At Chatime Binjai Supermall,” *JITE*, vol. 8, Mar. 2025, doi: 10.31289/jite.v8i3Sp.14431.
- [6] R. Hidayat *et al.*, “Implementasi Algoritma Random Forest Regression Untuk Memprediksi Penjualan Produksi di Supermarket,” *SIMKOM*, vol. 10, no. 1, pp. 101–109, Jan. 2025, doi: 10.51717/simkom.v10i1.703.
- [7] Y. Marilaeta Nurak, S. Wahyu Iriananda, and F. Marisa, “Prediksi Penjualan Warung Kopi Ol Menggunakan Metode Random Forest dan XGBoost,” 2025. [Online]. Available: <https://www.kaggle.com/datasets/keremkarayaz/coff>
- [8] M. A. Pandu W, R. E. Saputro, P. Purwadi, and U. A. Rohmah, “Analisis Tingkat Akurasi Metode Naive Bayes dan Random Forest dalam Prediksi Penjualan Emas,” *Jurnal Pendidikan dan Teknologi Indonesia*, vol. 5, no. 7, pp. 1809–1821, Jul. 2025, doi: 10.52436/1.jpti.732.
- [9] 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.
- [10] A. Rachman Andhika, “Machine Learning dalam Pengembangan Perangkat Lunak,” *Integrative Perspectives of Social and Science Journal*, vol. 2, no. 1, p. 1211, 2025.
- [11] A. M. Shimaoka, R. C. Ferreira, and A. Goldman, “The evolution of CRISP-DM for Data Science: Methods, Processes and Frameworks,” *SBC Reviews on Computer Science*, vol. 4, no. 1, pp. 28–43, Oct. 2024, doi: 10.5753/reviews.2024.3757.