

PREDIKSI OMZET HARIAN WARKOP WARMINDOBLEH DENGAN ALGORITMA RANDOM FOREST REGRESSION

Rifky As-Syudais, Betha Nurina Sari, Adhi Rizal

Informatika, Universitas Singaperbangsa Karawang

Jl. HS.Ronggo Waluyo, Karawang, Jawa Barat, Indonesia

1910631170125@student.unsika.ac.id

ABSTRAK

Penelitian ini bertujuan membangun model prediksi omzet harian pada Warkop Warmindobleh menggunakan algoritma Random Forest Regression berbasis kerangka kerja CRISP-DM guna mengatasi ketidakakuratan perencanaan tenaga kerja akibat fluktuasi permintaan harian yang dipengaruhi faktor temporal, cuaca, dan event lokal. Data yang digunakan mencakup 358 hari transaksi periode 1 Januari hingga 31 Desember 2024 yang diintegrasikan dengan variabel eksternal cuaca (*is_raining* hasil kategorisasi ordinal), event di Jakarta International Velodrome (*has_event*), serta fitur time-series lag (*previous_day_revenue* dan *revenue_7day_avg*). Model dilatih dengan 80% data awal (Januari–September) dan diuji pada 20% data akhir (Oktober–Desember) menggunakan split kronologis, kemudian dibandingkan dengan Linear Regression dan ARIMA melalui metrik MAE, RMSE, R^2 Score, dan MAPE serta divalidasi dengan 5-fold Time-Series Cross-Validation. Hasil evaluasi menunjukkan bahwa model Random Forest Regression (konfigurasi default) memberikan performa terbaik secara relatif dengan nilai R^2 sebesar 0,228, MAE Rp 163.204, RMSE Rp 211.023, dan MAPE 18,12% pada data uji; analisis feature importance mengungkapkan *revenue_7day_avg* sebagai faktor paling dominan (79,14%), diikuti *previous_day_revenue* (9,31%). Meskipun akurasi model belum memenuhi kriteria keberhasilan absolut ($R^2 > 0,85$ dan error di bawah 5–10% rata-rata omzet), model ini terbukti paling efektif di antara metode yang diuji dan berpotensi mendukung estimasi operasional UMKM warkop secara lebih objektif.

Kata kunci : *Random Forest Regression, Prediksi Omzet Harian, CRISP-DM, UMKM Warkop, Feature Importance, Time-Series Cross-Validation*

1. PENDAHULUAN

Usaha mikro, kecil, dan menengah (UMKM) di sektor kuliner, khususnya warung kopi (warkop), merupakan salah satu penopang perekonomian masyarakat yang memiliki karakteristik operasional sangat dinamis. Permintaan pelanggan pada usaha ini sangat dipengaruhi oleh berbagai faktor, antara lain pola kunjungan harian, kondisi cuaca, keberadaan event lokal, serta hari dalam minggu. Fluktuasi permintaan yang tidak terprediksi sering kali menyebabkan ketidaktepatan dalam perencanaan jumlah tenaga kerja, yang pada akhirnya berdampak pada penurunan kualitas pelayanan ketika kekurangan tenaga kerja atau peningkatan biaya operasional ketika kelebihan tenaga kerja. Tren konsumsi kopi di Indonesia yang terus meningkat pesat [2] semakin menegaskan pentingnya kemampuan pelaku usaha dalam mengantisipasi perubahan permintaan secara lebih akurat dan berbasis data.

Saat ini, sebagian besar warkop masih mengandalkan pengalaman subjektif pemilik dalam menentukan jumlah tenaga kerja harian tanpa didukung oleh analisis data historis yang sistematis. Kondisi ini menimbulkan kesenjangan antara kebutuhan bisnis dengan pemanfaatan teknologi analitik yang tersedia. Di sisi lain, berbagai penelitian telah menunjukkan bahwa penerapan algoritma *machine learning*, khususnya Random Forest Regression, mampu memberikan hasil prediksi penjualan yang lebih akurat pada sektor ritel dan UMKM [4], [5], [6], [7], [9]. Namun, penelitian-

penelitian tersebut umumnya belum mengintegrasikan secara simultan variabel cuaca, event lokal, dan fitur *lag* deret waktu dalam satu model prediksi omzet harian warkop. Selain itu, perbandingan langsung antara Random Forest Regression dengan metode klasik seperti Linear Regression dan ARIMA pada dataset yang sama juga masih jarang dilakukan.

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk membangun model prediksi omzet harian Warkop Warmindobleh menggunakan algoritma Random Forest Regression berbasis kerangka kerja CRISP-DM. Secara spesifik, tujuan penelitian ini adalah (1) mengembangkan model prediksi yang mengintegrasikan data transaksi harian, variabel cuaca, event lokal di Jakarta International Velodrome, serta fitur *time-series lag*; (2) mengevaluasi dan membandingkan performa model Random Forest Regression dengan Linear Regression dan ARIMA menggunakan metrik MAE, RMSE, R^2 Score, dan MAPE; serta (3) mengidentifikasi faktor-faktor yang paling berpengaruh terhadap prediksi omzet melalui analisis *feature importance*.

Untuk mencapai tujuan tersebut, penelitian ini direncanakan menggunakan pendekatan data mining dengan kerangka kerja CRISP-DM yang terdiri dari enam tahapan iteratif. Data yang digunakan mencakup transaksi harian periode 1 Januari hingga 31 Desember 2024 yang diintegrasikan dengan data cuaca historis dari Open-Meteo API dan jadwal event lokal. Model akan dilatih menggunakan 80% data awal dan diuji pada 20% data akhir dengan skema *chronological*

split. Evaluasi dilakukan secara komparatif terhadap tiga model, dilengkapi dengan validasi *5-fold time-series cross-validation* serta analisis *feature importance* untuk memberikan interpretasi bisnis yang lebih mendalam. Hasil penelitian diharapkan dapat memberikan kontribusi berupa model prediksi yang aplikatif serta rekomendasi operasional bagi pengelola warkop dalam meningkatkan efisiensi perencanaan tenaga kerja dan pengelolaan usaha secara keseluruhan.

2. TINJAUAN PUSTAKA

Prediksi penjualan dan omzet harian merupakan aspek fundamental dalam pengelolaan operasional UMKM sektor kuliner, khususnya warung kopi (warkop). Fluktuasi permintaan yang tidak terduga sering kali menyebabkan ketidaktepatan dalam perencanaan stok dan tenaga kerja, sehingga berdampak langsung pada efisiensi biaya dan kepuasan pelanggan [2]. Penelitian [2] menyoroti bahwa tren konsumsi kopi di Indonesia terus meningkat pesat, menciptakan kebutuhan mendesak bagi pelaku usaha untuk memiliki sistem prediksi yang andal. Beberapa studi terdahulu telah menunjukkan efektivitas machine learning dalam konteks ini. Penelitian [7] berhasil menerapkan Random Forest Regression untuk memprediksi pendapatan penjualan di Indomaret dengan akurasi sangat tinggi. Penelitian [6] melaporkan hasil serupa pada supermarket, di mana model Random Forest mampu menghasilkan prediksi penjualan dengan Mean Absolute Error yang rendah. Penelitian [4] dan [5] juga membuktikan bahwa penerapan Random Forest pada data penjualan produk UMKM dapat meningkatkan akurasi prediksi sekaligus mendukung pengelolaan persediaan yang lebih efisien. Penelitian [9] secara spesifik mengembangkan model prediksi penjualan harian usaha minuman menggunakan data Point of Sales (POS) dan mencapai tingkat akurasi yang memuaskan.

2.1. Pengaruh Faktor Eksternal Terhadap Omzet

Faktor eksternal seperti cuaca dan keberadaan event lokal terbukti memiliki pengaruh signifikan terhadap volume penjualan di industri makanan dan minuman. Penelitian [11] menemukan bahwa kondisi cuaca, terutama curah hujan dan suhu lingkungan, memengaruhi preferensi konsumen dalam memilih menu serta jumlah kunjungan ke restoran. Penelitian [12] semakin memperkuat temuan tersebut dengan menunjukkan bahwa dampak cuaca terhadap permintaan konsumen di sektor jasa bersifat kontekstual dan dapat bervariasi antar jenis usaha. Di samping faktor cuaca, keberadaan event juga menjadi variabel penting. Penelitian [10] menekankan bahwa pengelolaan event pada kegiatan usaha makanan perlu dilakukan secara terencana karena dapat meningkatkan kunjungan pelanggan secara substansial, meskipun juga berpotensi menimbulkan risiko operasional jika tidak diantisipasi dengan baik..

2.2. Random Forest Regression

Random Forest Regression merupakan salah satu algoritma ensemble learning yang paling populer dalam tugas prediksi numerik. Penelitian [3] memperkenalkan metode ini sebagai pengembangan dari pohon keputusan tunggal dengan menggabungkan sejumlah besar pohon yang dilatih secara independen menggunakan teknik bootstrap aggregating dan random feature selection. Keunggulan utama Random Forest terletak pada kemampuannya menangani hubungan non-linear, ketahanan terhadap outlier, serta kemampuan memberikan informasi feature importance yang berguna untuk interpretasi bisnis. Keunggulan ini telah dimanfaatkan secara luas dalam berbagai penelitian prediksi penjualan. Penelitian [4], [5], [6], [7], dan [9] secara konsisten melaporkan bahwa Random Forest mampu memberikan performa prediksi yang lebih baik dibandingkan metode regresi linier pada data penjualan yang bersifat fluktuatif. Penelitian [1] juga menunjukkan bahwa Random Forest tetap kompetitif ketika dibandingkan dengan Gradient Boosting dan Ridge Regression pada data deret waktu.

2.3. Metode Pembandingan

Sebagai pembandingan, penelitian ini menggunakan Linear Regression dan ARIMA. Linear Regression merupakan metode statistik yang mengasumsikan hubungan linear antara variabel independen dan dependen, serta dikenal memiliki interpretasi yang baik dan perhitungan yang relatif sederhana. Namun, metode ini memiliki keterbatasan dalam menangkap hubungan non-linear. Sementara itu, ARIMA merupakan metode time series forecasting yang memanfaatkan komponen autoregressive, integrated, and moving average dalam memodelkan data deret waktu. ARIMA cukup baik untuk data runtut waktu, tetapi membutuhkan asumsi tertentu seperti kestasioneran data dan kurang fleksibel dalam menangani banyak variabel eksternal secara bersamaan. Penggunaan kedua metode pembandingan ini bertujuan untuk menilai apakah Random Forest Regression benar-benar memberikan performa yang lebih baik pada kasus prediksi omzet harian.

2.4. Perbandingan Metode

Berbagai penelitian terdahulu telah melakukan perbandingan antar model prediksi untuk menentukan pendekatan yang paling sesuai dengan karakteristik data. Penelitian [1] membandingkan Random Forest Regression dengan Ridge Regression dan Gradient Boosting pada data curah hujan harian menggunakan *time series cross-validation*. Hasilnya menunjukkan bahwa Random Forest mampu memberikan performa kompetitif, meskipun tidak selalu menjadi model terbaik pada setiap fold validasi. Sementara itu, penelitian [9] menerapkan Random Forest Regression untuk prediksi penjualan harian usaha minuman tanpa melakukan perbandingan eksplisit dengan model lain,

namun mencatat nilai *Mean Absolute Error* yang sangat rendah (2,85%).

Penelitian [7] dan [6] juga menggunakan Random Forest Regression sebagai model utama pada data penjualan ritel dan supermarket dengan hasil akurasi yang sangat tinggi (R^2 mendekati 1,0). Namun, kedua studi tersebut tidak membandingkan Random Forest dengan metode deret waktu klasik seperti ARIMA. Di sisi lain, penelitian [8] memilih algoritma C4.5 dalam kerangka CRISP-DM untuk prediksi pendapatan, tetapi tidak melakukan evaluasi terhadap model *ensemble* seperti Random Forest.

Secara umum, mayoritas penelitian terdahulu lebih banyak berfokus pada penerapan Random Forest Regression tanpa melakukan perbandingan komprehensif dengan metode regresi linier dan model deret waktu (ARIMA). Kesenjangan ini menjadi dasar bagi penelitian saat ini untuk melakukan evaluasi komparatif antara Random Forest Regression, Linear Regression, dan ARIMA pada dataset omzet harian warkop yang sama.

Tabel 1. Perbandingan metode prediksi

Aspek	Linear Regression	ARIMA	Random Forest Regression
Hubungan linear	Baik	Baik	Sangat baik
Hubungan non-linear	Lemah	Lemah	Sangat baik
Penanganan outlier	Lemah	Lemah	Sangat baik
Banyak variabel	Moderate	Moderate	Sangat baik
Interpretabilitas	Tinggi	Tinggi	Moderate
Akurasi umum	Moderate	Moderate	Sangat baik

2.5. CRISP-DM

Dalam pengembangan model data mining yang sistematis, kerangka kerja CRISP-DM (Cross-Industry Standard Process for Data Mining) sering digunakan sebagai panduan standar. Ruswanti et al. [8] mendemonstrasikan penerapan CRISP-DM dalam proyek prediksi pendapatan menggunakan algoritma C4.5 dan membuktikan bahwa pendekatan terstruktur ini mampu meningkatkan kualitas data preparation, pemodelan, serta evaluasi model secara keseluruhan. Kerangka ini terdiri dari enam tahapan iteratif, yaitu Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, dan Deployment, yang sangat relevan untuk diterapkan pada konteks prediksi omzet harian warkop.

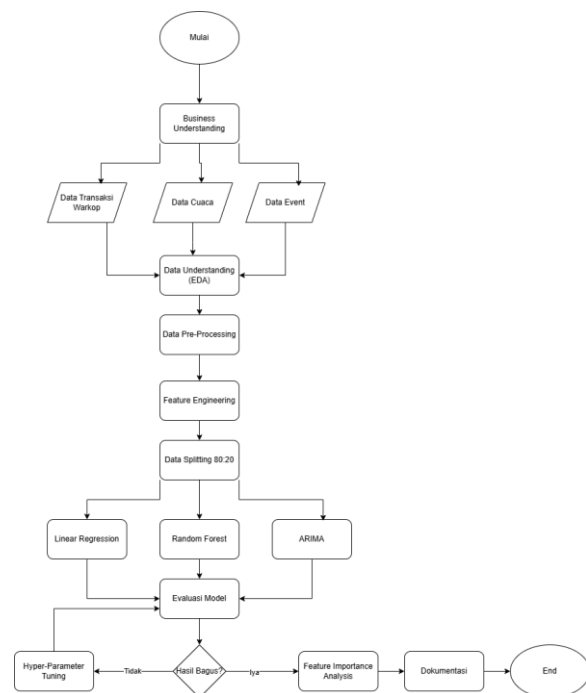
2.6. Penelitian Terdahulu

Berdasarkan tinjauan terhadap penelitian-penelitian sebelumnya, terlihat bahwa penerapan Random Forest Regression telah banyak dilakukan pada sektor ritel dan UMKM [4], [5], [6], [7], [9]. Namun, sebagian besar studi tersebut belum mengintegrasikan secara komprehensif variabel cuaca,

event lokal, dan fitur lag deret waktu dalam satu model prediksi omzet harian warkop. Selain itu, perbandingan langsung antara Random Forest dengan metode klasik seperti Linear Regression dan ARIMA pada dataset yang sama masih jarang dilakukan, khususnya pada konteks usaha warkop di Indonesia. Penelitian ini berusaha mengisi kesenjangan tersebut dengan membangun model prediksi omzet harian Warkop Warmindobleh menggunakan Random Forest Regression berbasis kerangka CRISP-DM, mengintegrasikan fitur temporal, cuaca, event, serta time-series lag, serta melakukan evaluasi komparatif yang sistematis terhadap model pembandingan.

3. METODE PENELITIAN

Penelitian ini menggunakan pendekatan *data mining* dengan kerangka kerja CRISP-DM untuk membangun model prediksi omzet harian Warkop Warmindobleh. Metode ini dipilih karena menyediakan alur kerja yang sistematis, mulai dari pemahaman masalah bisnis, pemahaman data, persiapan data, pemodelan, evaluasi, hingga penyusunan rekomendasi implementasi. Dalam penelitian ini, fokus utama diarahkan pada pembangunan model prediksi berbasis data historis transaksi yang diperkaya dengan variabel eksternal berupa cuaca dan event di sekitar lokasi usaha.



Gambar 1. Rancangan penelitian

Diagram tersebut menggambarkan hubungan antara sumber data transaksi, data cuaca, dan data event yang kemudian diproses melalui tahapan *data understanding*, *data preprocessing*, *feature engineering*, *data splitting*, pemodelan, evaluasi, dan analisis *feature importance*. Pada versi artikel, gambar ini cukup ditampilkan satu kali sebagai representasi metodologi secara keseluruhan.

3.1. Sumber dan Karakteristik Data

Data yang digunakan dalam penelitian mencakup periode 1 Januari 2024 sampai 31 Desember 2024. Untuk meningkatkan kualitas prediksi, penelitian ini mengintegrasikan tiga sumber data. Pertama, data primer internal, yaitu data transaksi penjualan harian atau omzet yang diekspor dari sistem pencatatan Warkop Warmindobleh. Kedua, data sekunder event, yaitu jadwal acara di Jakarta International Velodrome yang dikumpulkan melalui kurasi manual dari arsip berita digital dan media sosial resmi penyelenggara. Ketiga, data sekunder cuaca, yaitu data historis cuaca harian yang diperoleh dari Open-Meteo API berdasarkan titik koordinat lokasi usaha. Integrasi ketiga sumber data ini dilakukan agar model tidak hanya belajar dari pola penjualan masa lalu, tetapi juga mempertimbangkan faktor eksternal yang relevan.

3.2. Pra-pemrosesan Data dan Rekayasa Fitur

Tahap pra-pemrosesan dilakukan untuk memastikan data berada dalam kondisi layak sebelum digunakan pada proses pemodelan. Kegiatan utama pada tahap ini meliputi verifikasi dan penghapusan

duplikasi, penanganan missing values, pemeriksaan konsistensi format data, serta pencatatan outlier yang mungkin muncul pada data transaksi harian. Setelah proses pembersihan selesai, dilakukan rekayasa fitur untuk menghasilkan variabel yang lebih informatif bagi model. Fitur-fitur yang dibangun mencakup aspek temporal, seperti `day_of_week`, `is_weekend`, dan `day_of_month`, serta fitur kontekstual seperti `is_holiday`, `has_event`, dan `is_raining`. Selain itu, digunakan juga fitur berbasis time series berupa `previous_day_revenue` dan `revenue_7day_avg` untuk menangkap autokorelasi harian dan tren mingguan omzet. Khusus untuk data cuaca, kode cuaca mentah dari Open-Meteo ditransformasikan ke dalam kategori ordinal agar lebih mudah dipahami model. Kategori tersebut dibagi menjadi kondisi normal, hujan ringan, dan hujan berat berdasarkan logika bisnis yang disesuaikan dengan perilaku pelanggan pada usaha kuliner. Langkah ini dilakukan karena hujan ringan diasumsikan dapat meningkatkan pembelian menu hangat, sedangkan hujan berat cenderung menurunkan mobilitas pelanggan.

Tabel 2. Variabel prediktor

Variabel	Tipe data	Deskripsi
<code>day_of_week</code>	Numerik	Hari dalam seminggu untuk menangkap pola mingguan
<code>is_weekend</code>	Biner	Penanda akhir pekan
<code>day_of_month</code>	Numerik	Menangkap pola tanggal muda dan tanggal tua
<code>is_holiday</code>	Biner	Penanda hari libur nasional
<code>has_event</code>	Biner	Penanda adanya event di sekitar lokasi
<code>is_raining</code>	Ordinal	Representasi kondisi cuaca
<code>previous_day_revenue</code>	Numerik	Omzet hari sebelumnya
<code>revenue_7day_avg</code>	Numerik	Rata-rata omzet 7 hari sebelumnya

3.3. Transformasi Data dan Pembagian Data

Setelah fitur dibentuk, seluruh variabel prediktor digabungkan menjadi matriks fitur X , sedangkan omzet harian dijadikan vektor target y . Penelitian ini menggunakan pendekatan time-series split dengan komposisi 80% data training dan 20% data testing untuk menjaga urutan waktu dan menghindari data leakage. Berdasarkan pembagian tersebut, data training mencakup 293 hari, sedangkan data testing mencakup 72 hari. Pendekatan ini dipilih karena data penelitian berbentuk deret waktu, sehingga urutan kronologis perlu dipertahankan agar evaluasi lebih realistis.

3.4. Pemodelan

Model utama yang digunakan adalah Random Forest Regression dengan parameter dasar seperti `n_estimators = 100`, `max_depth = None`, `min_samples_split = 2`, `min_samples_leaf = 1`, dan `random_state = 42`. Selain model utama, penelitian ini juga menggunakan Linear Regression dan ARIMA sebagai model pembanding. Ketiga model dibangun menggunakan dataset yang sama agar hasil evaluasi bersifat objektif. Random Forest dipilih sebagai model utama karena memiliki kemampuan menangani hubungan non-linear dan menyediakan analisis feature

importance untuk melihat kontribusi masing-masing variabel terhadap hasil prediksi.

3.5. Evaluasi Model

Evaluasi model dilakukan menggunakan tiga metrik utama, yaitu Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), dan R-Squared (R^2). RMSE digunakan untuk mengukur besar kesalahan prediksi dengan sensitivitas lebih tinggi terhadap error besar, MAE digunakan untuk melihat rata-rata deviasi absolut hasil prediksi terhadap nilai aktual, dan R^2 digunakan untuk melihat proporsi variasi omzet yang dapat dijelaskan oleh model. Selain evaluasi pada data uji, penelitian ini juga menggunakan 5-fold time-series cross-validation sebagai pengujian tambahan untuk melihat kestabilan performa model.

4. HASIL DAN PEMBAHASAN

4.1. Hasil

Hasil analisis menunjukkan bahwa data omzet harian Warkop Warmindobleh sepanjang tahun 2024 memiliki fluktuasi yang cukup tinggi, sehingga layak digunakan sebagai dasar pembangunan model prediksi. Berdasarkan hasil analisis omzet, dataset yang digunakan mencakup 358 hari operasional

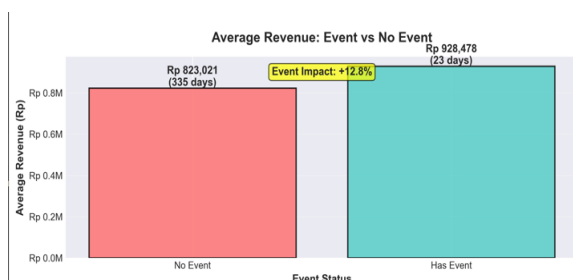
dengan total omzet sebesar Rp 296.947.204. Rata-rata omzet harian tercatat sebesar Rp 829.797, median sebesar Rp 808.000, dan standar deviasi sebesar Rp 268.000. Selain itu, omzet minimum tercatat sebesar Rp 146.000, sedangkan omzet maksimum mencapai Rp 1.709.000. Nilai rata-rata yang sedikit lebih tinggi daripada median menunjukkan bahwa distribusi omzet cenderung sedikit condong ke kanan (*right-skewed*), yang berarti terdapat beberapa hari dengan omzet sangat tinggi yang menarik rata-rata ke atas. Kondisi ini memperlihatkan bahwa pola penjualan Warmindobleh memang memiliki variasi yang cukup besar antarmhari.



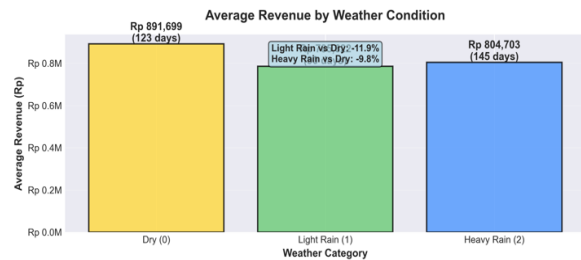
Gambar 2. Hasil analisis omzet

Visualisasi pada Gambar 2 memperlihatkan bahwa mayoritas hari operasional mencatatkan omzet pada rentang Rp 700.000 hingga Rp 900.000, meskipun terdapat pencapaian signifikan pada hari-hari tertentu yang melampaui angka Rp 1.500.000. Adanya lonjakan tersebut mengindikasikan bahwa karakteristik omzet harian memiliki tingkat fluktuasi yang dipengaruhi oleh aspek temporal serta variabel eksternal seperti keberadaan event lokal. Fenomena ini memperkuat urgensi penggunaan pendekatan *machine learning* guna memodelkan pola fluktuatif tersebut melalui analisis data yang lebih sistematis.

Hasil analisis rata-rata omzet berdasarkan kategori hari menunjukkan bahwa omzet weekday justru lebih tinggi dibandingkan weekend. Rata-rata omzet pada weekday tercatat sebesar Rp 855.791 pada sekitar 254 hari, sedangkan rata-rata omzet weekend sebesar Rp 762.730 pada sekitar 104 hari. Dengan demikian, omzet pada akhir pekan tercatat sekitar 10,9% lebih rendah dibandingkan hari kerja. Hasil ini menunjukkan bahwa asumsi awal bahwa akhir pekan selalu lebih ramai tidak terbukti pada kasus Warmindobleh. Variabel *is_weekend* tetap penting digunakan dalam model, tetapi arah pengaruhnya cenderung negatif terhadap omzet.



Gambar 3. Rata-rata omzet faktor event



Gambar 4. Rata-rata omzet faktor cuaca

Analisis faktor eksternal menunjukkan bahwa perbedaan rata-rata omzet juga terlihat pada hari dengan event dan tanpa event, serta pada kondisi cuaca yang berbeda. Variabel event digunakan untuk menandai adanya kegiatan besar di sekitar Jakarta International Velodrome, sedangkan variabel cuaca ditransformasikan menjadi kategori ordinal agar lebih mudah dipelajari model. Temuan ini memperlihatkan bahwa selain pola historis omzet, faktor eksternal juga memiliki kontribusi dalam menjelaskan variasi pendapatan harian, walaupun tingkat pengaruhnya tidak sebesar fitur historis.

Setelah tahap eksplorasi data, dilakukan pembangunan tiga model prediksi, yaitu Random Forest Regression sebagai model utama, serta Linear Regression dan ARIMA sebagai model pembanding. Evaluasi dilakukan pada data uji menggunakan metrik MAE, RMSE, R^2 , dan MAPE. Hasil evaluasi menunjukkan bahwa Random Forest Regression (default) memberikan performa terbaik secara relatif dibandingkan model pembanding, dengan nilai R^2 sebesar 0,228, MAE sebesar Rp 163.204, dan MAPE sebesar 18,12% pada data uji. Hasil ini menunjukkan bahwa Random Forest merupakan model yang paling mampu mengikuti pola omzet harian dibandingkan dua metode lainnya.

Tabel 3. Hasil evaluasi model

Metrik Evaluasi	ARIMA (1,1,1)	Linear Regression	Random Forest (Tuned)	Random Forest (Default)
MAE (Rupiah)	Rp 197.697	Rp 162.980	Rp 165.540	Rp 163.204
RMSE (Rupiah)	Rp 240.563	Rp 210.905	Rp 214.252	Rp 211.023
R^2 Score	-0,003	0,229	0,204	0,228
MAPE (%)	22,08%	18,53%	18,62%	18,12%

Dari Tabel 3 dapat dilihat bahwa model Random Forest Regression memberikan hasil evaluasi paling baik dibandingkan model pembanding. Hal ini sejalan dengan karakteristik algoritma Random Forest yang mampu menangani hubungan non-linear dan interaksi antarfitur dengan lebih baik dibandingkan regresi linear maupun metode deret waktu yang lebih sederhana.

Tabel 4. Hasil analisis fitur

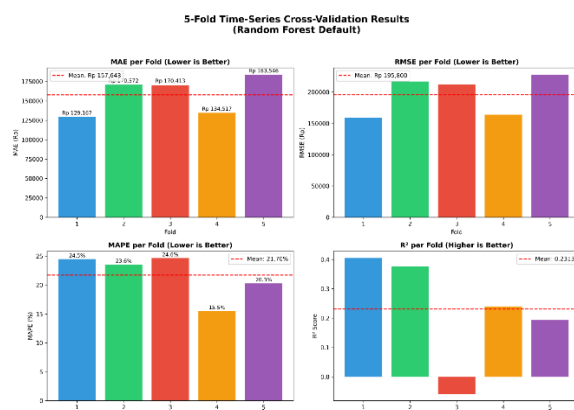
Rank	Feature	Tingkat Kepentingan (%)	Kumulatif
1	revenue_7day_avg	79,14%	79,14%
2	previous_day_revenue	9,31%	88,46%
3	day_of_month	5,38%	93,83%
4	day_of_week	4,60%	98,44%
5	is_raining	0,93%	99,37%
6	is_weekend	0,55%	99,92%
7	is_holiday	0,06%	99,97%
8	has_event	0,02%	100,00%

Tabel 5. Hasil uji cross-validation

Fold	Periode Latih	Periode Validasi	Train	Val	MAE (Rp)	RMSE (Rp)	R ²	MAPE (%)
1	01 Jan – 03 Mar	04 Mar – 08 Mei	63	59	129,167	158,679	0,405	24,47%
2	01 Jan – 08 Mei	09 Mei – 06 Jul	122	59	170,572	217,042	0,376	23,56%
3	01 Jan – 06 Jul	07 Jul – 03 Sep	181	59	170,413	211,995	-0,058	24,64%
4	01 Jan – 03 Sep	04 Sep – 01 Nov	240	59	134,517	163,758	0,240	15,51%
5	01 Jan – 01 Nov	02 Nov – 31 Des	299	59	183,546	227,524	0,194	20,29%

Tabel 6. Rata-rata hasil cross-validation

Metrik	Rata-Rata	Standar Deviasi
MAE	Rp 157,643	Rp 23,621
RMSE	Rp 195,800	Rp 31,380
R ²	0,231	0,168
MAPE	21,70%	3,62%



Gambar 5. Visualisasi hasil cross-validation

Hasil 5-fold time-series cross-validation menunjukkan bahwa performa model berada pada tingkat yang cukup konsisten di beberapa potongan waktu. Rata-rata hasil cross-validation menunjukkan nilai MAE sebesar Rp 157.643, RMSE sebesar Rp 195.800, R² sebesar 0,231, dan MAPE sebesar 21,70%. Standar deviasi MAPE sebesar 3,62% menunjukkan bahwa variasi performa antar-fold masih berada dalam kisaran moderat. Dengan demikian, model tidak menunjukkan kegagalan ekstrem pada sebagian besar skenario pengujian, walaupun akurasi masih belum tergolong sangat tinggi.

4.2. Pembahasan

Hasil penelitian menunjukkan bahwa karakteristik data omzet harian Warmindobleh bersifat fluktuatif dan tidak sepenuhnya stabil sepanjang tahun.

Analisis feature importance menunjukkan bahwa revenue_7day_avg merupakan fitur paling dominan dengan kontribusi sebesar 79,14%. Hasil ini menunjukkan bahwa rata-rata omzet tujuh hari sebelumnya menjadi faktor utama yang memengaruhi prediksi omzet hari berikutnya. Dibandingkan faktor eksternal seperti event dan cuaca, fitur historis ini jauh lebih kuat dalam menjelaskan pola omzet harian. Temuan tersebut menunjukkan bahwa pola pendapatan jangka pendek memiliki peran sentral dalam pembentukan model prediksi Warmindobleh.

Kondisi ini terlihat dari besarnya standar deviasi omzet dan adanya beberapa hari dengan lonjakan pendapatan yang sangat tinggi. Pola seperti ini menyebabkan model prediksi harus mampu menangkap hubungan yang tidak sepenuhnya linear. Dalam konteks ini, performa Random Forest Regression yang lebih baik dibandingkan Linear Regression dan ARIMA menunjukkan bahwa pendekatan berbasis ensemble lebih sesuai untuk memodelkan variasi omzet harian yang kompleks.

Jika dibandingkan dengan gaya pembahasan pada artikel contoh JATI, hasil penelitian ini juga menunjukkan pola yang serupa, yaitu adanya perbedaan antara performa terbaik pada data uji dan konsistensi model pada validasi temporal. Pada artikel contoh, model terbaik pada data uji dan model paling stabil pada cross-validation tidak selalu identik. Pola serupa juga terlihat pada penelitian ini, di mana evaluasi tidak cukup hanya mengandalkan satu metrik, tetapi perlu melihat kestabilan model pada beberapa rentang waktu. Pendekatan ini penting karena data deret waktu sangat sensitif terhadap perubahan pola musiman dan konteks operasional.

Dari sisi interpretasi bisnis, dominasi fitur revenue_7day_avg menunjukkan bahwa omzet harian Warmindobleh lebih banyak dipengaruhi oleh pola historis jangka pendek dibandingkan faktor eksternal. Artinya, perilaku omzet cenderung mengikuti tren mingguan yang sudah terbentuk sebelumnya. Meskipun demikian, variabel seperti event, cuaca, dan weekend tetap berguna sebagai variabel penjelas tambahan karena membantu model mengenali kondisi khusus pada hari-hari tertentu. Temuan ini relevan dengan studi terdahulu yang menunjukkan bahwa kombinasi fitur historis dan fitur kontekstual dapat meningkatkan kualitas prediksi penjualan.

Walaupun Random Forest Regression menjadi model terbaik secara relatif, hasil implementasi menunjukkan bahwa performa model masih belum

memenuhi target awal penelitian. Berdasarkan perbandingan terhadap target proposal, nilai MAE sebesar Rp 163.204 masih melebihi ambang target, nilai RMSE sebesar Rp 211.023 juga masih lebih tinggi dari batas yang diharapkan, dan nilai R^2 sebesar 0,228 masih jauh di bawah target 0,85. Dengan demikian, model yang dibangun dalam penelitian ini belum dapat dikatakan sangat akurat secara absolut, tetapi sudah cukup menunjukkan bahwa pendekatan Random Forest lebih efektif dibandingkan metode pembandingan dalam konteks data yang tersedia.

Keterbatasan hasil ini dapat dijelaskan oleh beberapa faktor. Pertama, jumlah fitur yang digunakan masih terbatas, sehingga belum seluruh variasi omzet dapat dijelaskan oleh model. Kedua, data hanya mencakup satu tahun operasional, sehingga pola jangka panjang dan perubahan musiman tahunan belum dapat ditangkap secara lebih kuat. Ketiga, terdapat kemungkinan adanya faktor lain yang berpengaruh tetapi belum tercatat dalam dataset, seperti detail promosi, kondisi persaingan, perilaku pelanggan, atau dinamika operasional internal. Oleh karena itu, pengembangan penelitian lanjutan dapat diarahkan pada penambahan periode data, penambahan fitur eksternal yang lebih kaya, serta eksplorasi model hybrid atau tuning parameter yang lebih mendalam agar akurasi prediksi dapat ditingkatkan.

Secara keseluruhan, penelitian ini menunjukkan bahwa Random Forest Regression merupakan pendekatan yang paling layak digunakan untuk prediksi omzet harian Warkop Warmindobleh pada dataset yang tersedia. Model ini belum sempurna, tetapi telah mampu memberikan gambaran prediktif yang lebih baik dibandingkan regresi linear dan ARIMA. Dengan demikian, hasil penelitian ini dapat dijadikan dasar awal untuk mendukung pengambilan keputusan operasional berbasis data pada usaha warkop, terutama dalam membaca kecenderungan omzet harian dan merespons perubahan pola penjualan secara lebih terukur.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian, dapat disimpulkan bahwa model Random Forest Regression merupakan metode yang paling efektif secara relatif untuk memprediksi omzet harian Warkop Warmindobleh dibandingkan Linear Regression dan ARIMA pada dataset yang digunakan. Model ini mampu memberikan hasil evaluasi terbaik pada data uji dengan nilai R^2 sebesar 0,228, MAE sebesar Rp 163.204, dan MAPE sebesar 18,12%, sehingga lebih baik dalam mengikuti pola omzet harian yang fluktuatif. Hasil analisis juga menunjukkan bahwa fitur `revenue_7day_avg` merupakan variabel yang paling berpengaruh dengan kontribusi sebesar 79,14%, yang menandakan bahwa pola historis omzet jangka pendek menjadi faktor utama dalam proses prediksi. Meskipun demikian, performa model masih belum memenuhi target akurasi absolut yang diharapkan,

sehingga penelitian ini menunjukkan bahwa pendekatan *Random Forest* sudah layak dijadikan dasar prediksi awal, tetapi masih memerlukan pengembangan lebih lanjut. Oleh karena itu, penelitian selanjutnya disarankan menambahkan periode data yang lebih panjang, memperkaya variabel eksternal seperti promosi, perilaku pelanggan, dan faktor lingkungan lainnya, serta menguji strategi *hyperparameter tuning* atau model hybrid agar akurasi prediksi omzet harian dapat ditingkatkan.

DAFTAR PUSTAKA

- [1] M. Alvines et al., "Komparasi ridge regression, random forest, dan gradient boosting untuk prediksi curah hujan harian di Sumatra Selatan berbasis time series cross-validation," JATI, vol. 13, no. 5, pp. 145–162, 2025, doi: 10.36040/jati.v13i5.13915.
- [2] Arlinah, M. Rizki, and A. Praditya, "Tren pertumbuhan konsumsi kopi dan implikasinya terhadap pengembangan industri kopi Indonesia," Jurnal Agribisnis Indonesia, vol. 11, no. 3, pp. 234–248, 2023.
- [3] L. Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.
- [4] Darmanto and E. S. Barus, "Implementasi metode random forest untuk memprediksi penjualan produk," Jurnal TEKINKOM, vol. 7, no. 2, pp. 591–603, 2024, doi: 10.37600/tekinkom.v7i2.1510.
- [5] Efendi and Zyen, "Penerapan algoritma random forest untuk prediksi penjualan dan sistem persediaan produk," Jurnal Resolusi, vol. 5, no. 1, pp. 12–20, 2024, doi: 10.30865/resolusi.v5i1.2149.
- [6] R. Hidayat et al., "Implementasi Algoritma Random Forest Regression untuk memprediksi penjualan produksi di supermarket," Jurnal Sistem Informasi dan Sistem Komputer, vol. 10, no. 1, pp. 101–109, 2025, doi: 10.51717/simkom.v10i1.703.
- [7] M. A. Kurniawan, G. Zaidan Syauqi, M. Safriyanti, F. Ullul Azmie, and A. Setiawan, "Prediksi pendapatan penjualan di Indomaret menggunakan algoritma random forest regression," JSI (Jurnal Sistem Informasi) Universitas Suryadarma, vol. 12, no. 2, pp. 93–99, 2025.
- [8] D. Ruswanti, D. Susilo, and Riani, "Implementasi CRISP-DM pada data mining untuk melakukan prediksi pendapatan dengan algoritma C.45," GoINFORMATIK: Jurnal Informatika dan Teknologi Informasi, vol. 30, no. 1, pp. 111–125, 2024, doi: 10.36309/goi.v30i1.266.
- [9] R. Verdiyanto, D. Hartanti, and E. Purwanto, "Pengembangan aplikasi Point of Sales untuk prediksi penjualan harian usaha minuman menggunakan algoritma Random Forest

- Regression,” *Infotek: Jurnal Informatika dan Teknologi*, vol. 8, pp. 128–139, 2025, doi: 10.29408/jit.v8i1.28386.
- [10] A. Sari et al., “Event operation and management in risk assessment for outdoor food bazaar events,” *International Research Journal of Business Studies*, vol. 17, no. 2, pp. 163–170, 2024, doi: 10.21632/irjbs.17.2.163-170.
- [11] M. Bujisic, V. Bogicevic, and H. G. Parsa, “The effect of weather factors on restaurant sales,” *Journal of Foodservice Business Research*, vol. 20, no. 3, pp. 350–370, 2017.
- [12] W. Huang and J. Van Ossenbruggen, “Weather impact on consumer demand: Evidence from service industries,” *Journal of Retailing and Consumer Services*, vol. 50, pp. 20–29, 2019, doi: 10.1016/j.jretconser.2019.04.004.