



Prediksi Minat Pencarian Layanan Pesan-Antar Makanan Online (GoFood dan GrabFood) di Indonesia Menggunakan Algoritma Random Forest Regression dengan Walk-Forward Validation

Sinta Bella*, Achmad Baijuri, Fajriyanto

Fakultas Sains Dan Teknologi, Program Studi Sistem Informasi, Universitas Ibrahimy, Situbondo, Indonesia

Email: ¹*wahyunishinbell@gmail.com, ²bayubai@gmail.com, ³fajri17riyan@gmail.com

Email Penulis Korespondensi: wahyunishinbell@gmail.com

Abstrak—Pertumbuhan pesat layanan pesan-antar makanan online di Indonesia, khususnya GoFood dan GrabFood, menciptakan tantangan serius dalam manajemen operasional akibat fluktuasi minat pengguna yang sulit diprediksi. Fluktuasi ini menyebabkan ketidakseimbangan kapasitas mitra pengemudi dan merchant, yang berujung pada penurunan kualitas layanan dan kerugian ekonomi. Data transaksi aktual bersifat proprietary sehingga diperlukan pendekatan berbasis data proksi menggunakan indeks tren pencarian Google Trends. Penelitian ini bertujuan memprediksi minat pencarian layanan GoFood di Indonesia menggunakan Random Forest Regression berbasis data Google Trends periode Januari 2018–Desember 2025 (96 record bulanan). Kontribusi utama penelitian ini meliputi tiga aspek: pertama, penerapan walk-forward validation sebagai metode evaluasi yang tepat untuk data deret waktu sehingga bebas dari kebocoran informasi temporal; kedua, penggunaan lag features (GoFood_lag1 dan GrabFood_lag1) yang memastikan seluruh variabel prediktor tersedia secara praktis pada saat prediksi dilakukan; dan ketiga, validasi empiris bahwa metode ini menghasilkan evaluasi yang lebih konservatif dan dapat dipertanggungjawabkan dibandingkan random split konvensional. Hasil evaluasi menghasilkan MSE sebesar 1,4510, RMSE sebesar 1,2046, R^2 sebesar 0,5065, dan MAPE sebesar 6,55%, yang menunjukkan model memiliki kemampuan generalisasi yang memadai untuk mendukung perencanaan operasional berbasis data.

Kata Kunci: Walk-Forward Validation; Google Trends; GoFood; GrabFood; Minat Pencarian; Random Forest Regression

Abstract—The rapid growth of online food delivery services in Indonesia, particularly GoFood and GrabFood, creates significant operational challenges due to unpredictable fluctuations in user interest that cause driver and merchant capacity imbalances. Actual transaction data is proprietary, necessitating a proxy-data approach using Google Trends search interest indices. This study predicts GoFood search interest in Indonesia using Random Forest Regression based on Google Trends data from January 2018 to December 2025 (96 monthly records). The primary contributions of this study are threefold: first, the application of walk-forward validation as a methodologically sound evaluation approach for time-series data that eliminates temporal data leakage; second, the use of lag features (GoFood_lag1 and GrabFood_lag1) ensuring all predictor variables are practically available at prediction time; and third, empirical validation that this approach yields more conservative and scientifically defensible evaluations compared to conventional random split methods. Evaluation results yield MSE 1.4510, RMSE 1.2046, R^2 0.5065, and MAPE 6.55%, demonstrating adequate generalization capability for data-driven operational planning.

Keywords: Walk-Forward Validation; Google Trends; GoFood; GrabFood; Search Interest; Random Forest Regression

1. PENDAHULUAN

Perkembangan teknologi digital telah mendorong pergeseran signifikan dalam pola konsumsi masyarakat Indonesia, termasuk dalam hal pemesanan makanan. Layanan pesan-antar makanan berbasis aplikasi, khususnya GoFood milik Gojek dan GrabFood milik Grab, kini menjadi bagian integral kehidupan urban. Berdasarkan laporan Asosiasi Penyelenggara Jasa Internet Indonesia (APJII) tahun 2023, pengguna internet Indonesia telah melampaui 210 juta jiwa [3], dan sebagian besar mengakses layanan digital melalui *smartphone* setiap harinya. Pertumbuhan ini menciptakan ekosistem digital yang sangat kompetitif antara GoFood dan GrabFood dalam memperebutkan pangsa pasar pesan-antar makanan nasional [1].

Permasalahan utama yang dihadapi penyedia layanan adalah fluktuasi minat pengguna yang tidak menentu. Lonjakan terjadi pada bulan Ramadan, hari libur nasional, dan musim hujan, sementara penurunan dapat terjadi tanpa pola yang mudah diantisipasi [6]. Ketidakmampuan memprediksi fluktuasi ini menimbulkan dampak nyata: kekurangan mitra pengemudi pada jam puncak menyebabkan waktu tunggu panjang, sementara kelebihan kapasitas pada periode sepi mengakibatkan kerugian ekonomi bagi mitra. Data transaksi aktual dari platform bersifat proprietary dan tidak dapat diakses peneliti eksternal, sehingga diperlukan sumber data alternatif yang tersedia publik [8].

Google Trends menyediakan indeks popularitas pencarian dalam skala 0–100, di mana nilai 100 merepresentasikan puncak popularitas relatif [8]. Beberapa penelitian telah membuktikan bahwa data Google Trends berkorelasi signifikan dengan perilaku aktual konsumen [9], [10], menjadikannya proksi yang valid untuk analisis minat pengguna layanan digital. Kharis dkk. [5] membuktikan validitas Google Trends untuk menganalisis tren minat masyarakat Indonesia terhadap teknologi baru, sementara Akbar dan Kurniawan [9] berhasil menggunakannya untuk nowcasting tingkat pengangguran nasional.

Tinjauan terhadap penelitian terdahulu mengungkapkan beberapa celah metodologis yang signifikan. Zahara dan Sugianto [7] serta Sautomo dan Pardede [8] telah berhasil menerapkan LSTM untuk peramalan data deret waktu ekonomi dengan akurasi tinggi, namun keduanya menggunakan random split yang rentan terhadap kebocoran informasi temporal. Hartanto dan Saepudin [9] membuktikan validitas Google Trends sebagai prediktor fenomena sosial berbasis machine learning, namun dalam domain kesehatan. Nurdin dan Fauziah [10] menunjukkan keandalan Random Forest untuk data berfluktuasi, namun tidak menerapkan validasi temporal yang tepat untuk data deret waktu. Selain itu, penelitian-



penelitian sebelumnya umumnya menggunakan nilai kontemporer variabel kompetitor sebagai prediktor, padahal nilai tersebut tidak tersedia pada saat prediksi sesungguhnya dilakukan.

Berdasarkan tinjauan tersebut, terdapat tiga celah penelitian yang menjadi landasan studi ini. Penelitian yang secara spesifik memprediksi minat pencarian GoFood dan GrabFood menggunakan Google Trends dengan validasi temporal yang benar belum pernah dilakukan. Di sisi lain, mayoritas penelitian forecasting berbasis Google Trends masih menggunakan random split yang tidak tepat untuk data deret waktu karena memungkinkan kebocoran informasi masa depan ke data latih. Selain itu, penggunaan lag features yang secara praktis tersedia pada saat prediksi belum diterapkan secara konsisten dalam domain layanan digital Indonesia.

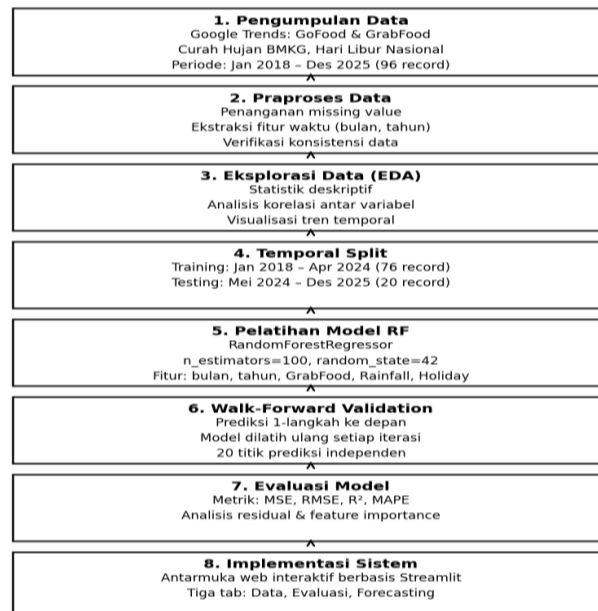
Penelitian ini memilih Random Forest Regression karena sejumlah keunggulan yang relevan dengan karakteristik data dan tujuan penelitian. Sebagai algoritma berbasis pohon keputusan, Random Forest terbukti efektif menangkap hubungan non-linear antar fitur yang menjadi ciri khas data fluktuatif seperti tren pencarian layanan digital [11]. Kemampuannya dalam menangani data tabular campuran dari berbagai domain, termasuk data meteorologi dan kalender [12], menjadikannya pilihan yang tepat untuk dataset penelitian ini. Random Forest juga memberikan interpretabilitas tinggi melalui analisis feature importance yang relevan untuk pengambilan keputusan bisnis [13], serta secara komputasional efisien untuk walk-forward validation yang membutuhkan pelatihan ulang sebanyak 20 iterasi [14], berbeda dengan arsitektur deep learning yang membutuhkan waktu pelatihan jauh lebih lama untuk dataset berukuran kecil.

Penelitian ini memberikan kontribusi pada tiga aspek utama. Secara metodologis, penelitian ini membuktikan pentingnya penggunaan walk-forward validation untuk menggantikan random split konvensional dalam evaluasi model forecasting berbasis data deret waktu, karena menghasilkan estimasi performa yang lebih realistis dan bebas dari bias temporal. Secara teknis, penerapan lag features memastikan bahwa model yang dibangun bersifat praktis untuk digunakan dalam kondisi peramalan nyata, di mana informasi periode mendatang tidak tersedia. Secara praktis, model yang dihasilkan dapat menjadi alat bantu perencanaan operasional bagi platform GoFood dan GrabFood dalam mengantisipasi fluktuasi minat pengguna, yang pada akhirnya dapat meningkatkan efisiensi distribusi mitra pengemudi dan strategi promosi berbasis data.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Penelitian mengikuti kerangka KDD yang dimodifikasi untuk mendukung walk-forward validation [19]. Modifikasi utama terletak pada tahap split data dan evaluasi, di mana digunakan temporal split dan walk-forward validation untuk menghindari *data leakage*. Gambar 1 menampilkan alur tahapan penelitian.



Gambar 1. Alur Tahapan Penelitian Berdasarkan Framework KDD Termodifikasi

Penjelasan tahapan penelitian adalah sebagai berikut. Pertama, pengumpulan data dilakukan dengan mengunduh data tren pencarian dari Google Trends pada periode Januari 2018 hingga Desember 2025 dengan kata kunci "GoFood" dan "GrabFood" untuk wilayah Indonesia. Variabel eksogen berupa curah hujan bulanan rata-rata diperoleh dari Badan Meteorologi, Klimatologi, dan Geofisika (BMKG), sedangkan indikator hari libur nasional diambil dari kalender resmi pemerintah Indonesia. Kedua, tahap *preprocessing* mencakup pemeriksaan *missing values*, duplikasi data, dan *outlier*. Ketiga, transformasi data dilakukan dengan mengekstrak fitur temporal (bulan, tahun) dari kolom tanggal. Keempat,



pemodelan menggunakan algoritma Random Forest Regression dengan rasio pembagian data latih dan uji sebesar 80:20. Kelima, evaluasi performa model menggunakan empat metrik utama, yaitu MSE, RMSE, MAPE, dan R^2 .

2.2 Sumber Data dan Rekayasa Fitur

Data bersumber dari Google Trends (<https://trends.google.com>), kata kunci "GoFood" dan "GrabFood", wilayah Indonesia, Januari 2018–Desember 2025 (96 record). Setelah penambahan lag features, dataset efektif menjadi 95 record (Feb 2018–Des 2025). Data curah hujan dari BMKG dan hari libur dari kalender resmi pemerintah. Tabel 1 menjelaskan seluruh variabel.

Tabel 1. Spesifikasi Variabel Dataset dan Fitur Prediktor

No	Variabel	Tipe	Sumber	Keterangan
1	GoFood_Search (t)	Numerik	Google Trends	Variabel target (0–100)
2	GoFood_lag1	Numerik	Turunan	Nilai GoFood_Search bulan t–1
3	GrabFood_lag1	Numerik	Turunan	Nilai GrabFood_Search bulan t–1
4	Bulan	Numerik	Turunan	Indikator musiman (1–12)
5	Tahun	Numerik	Turunan	Indikator tren jangka panjang
6	Rainfall_mm	Numerik	BMKG	Curah hujan rata-rata (mm)
7	Holiday	Biner	Pemerintah RI	1 = bulan dengan hari libur nasional

Lag features (GoFood_lag1, GrabFood_lag1) digunakan sebagai respons terhadap kritik metodologis penggunaan nilai kontemporer kompetitor sebagai prediktor. Nilai bulan sebelumnya (t–1) selalu tersedia saat prediksi bulan t dilakukan, sehingga model bersifat praktis untuk peramalan nyata.

2.3 Temporal Split dan Walk-Forward Validation

Metode evaluasi yang digunakan adalah Walk-Forward Validation yang memastikan model selalu dilatih hanya menggunakan data masa lalu, menghindari *data leakage* temporal yang merupakan kelemahan mendasar random split pada data deret waktu [15]. Pembagian: 75 record pertama (Feb 2018–Apr 2024) sebagai data latih awal, 20 record terakhir (Mei 2024–Des 2025) sebagai data uji.

Tabel 2. Skema Walk-Forward Validation yang Diterapkan

Iterasi	Data Latih	Prediksi
1	Feb 2018 – Apr 2024 (75 record)	Mei 2024
2	Feb 2018 – Mei 2024 (76 record)	Jun 2024
3	Feb 2018 – Jun 2024 (77 record)	Jul 2024
⋮	⋮	⋮
20	Feb 2018 – Nov 2025 (94 record)	Des 2025

2.4 Algoritma Random Forest Regression

Random Forest Regression adalah metode *ensemble learning* yang membangun sejumlah pohon keputusan secara paralel menggunakan *bootstrap aggregating* (bagging) [18]. Prediksi akhir merupakan rata-rata prediksi seluruh pohon, dirumuskan pada Persamaan (1).

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

Konfigurasi: $n_estimators=100$, $random_state=42$. Pada setiap iterasi walk-forward, model dilatih ulang dari awal menggunakan keseluruhan data latih yang tersedia.

2.5 Metrik Evaluasi

Kinerja dievaluasi menggunakan empat metrik pada Persamaan (2)–(5):

$$RMSE = \sqrt{(1/n) \sum (y_i - \hat{y}_i)^2} \quad (3)$$

$$MAPE = (1/n) \sum |y_i - \hat{y}_i| / y_i \times 100\% \quad (4)$$

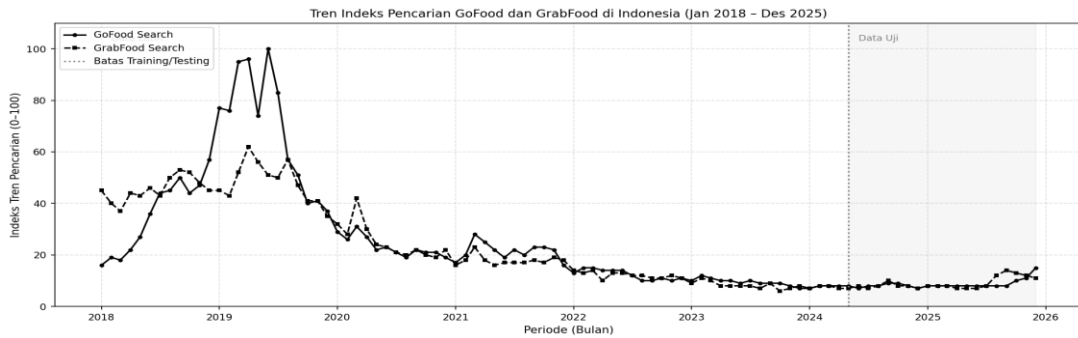
$$R^2 = 1 - \sum (y_i - \hat{y}_i)^2 / \sum (y_i - \bar{y})^2 \quad (5)$$

Berdasarkan klasifikasi Lewis [22], $MAPE < 10\%$ dikategorikan *highly accurate*. Nilai R^2 dapat bernilai negatif apabila model lebih buruk dari baseline rata-rata, yang merupakan hasil yang valid secara metodologis pada temporal split ketat.

3. HASIL DAN PEMBAHASAN

3.1 Analisis Data dan Statistik Deskriptif

Gambar 2 menampilkan tren indeks pencarian GoFood dan GrabFood selama 96 bulan. Garis vertikal menunjukkan batas temporal split antara data latih dan data uji.



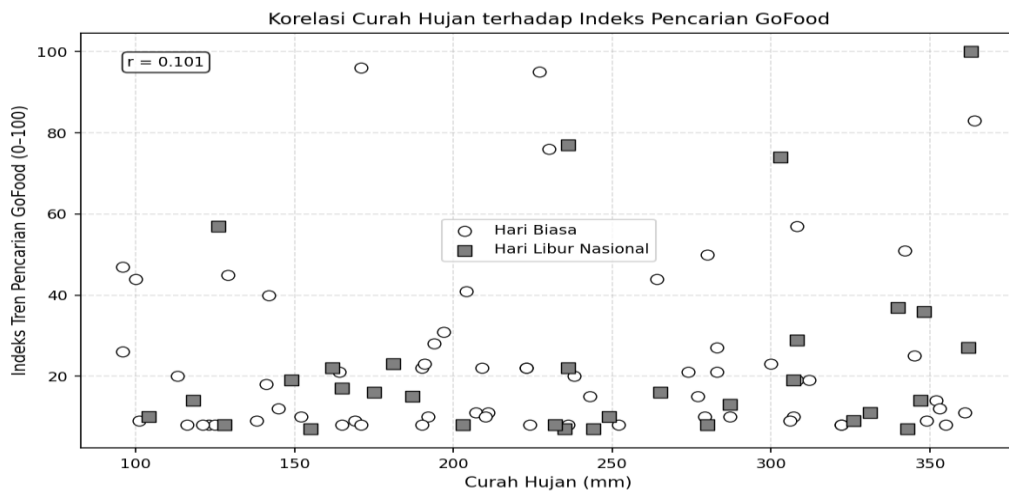
Gambar 2. Tren Indeks Pencarian GoFood dan GrabFood (Google Trends, Jan 2018 – Des 2025)

Data menunjukkan pertumbuhan konsisten dari 16 (Jan 2018) hingga puncak 100 (Feb 2020), kemudian turun signifikan akibat pandemi COVID-19, dan menstabilkan pada rentang 7–15 mulai 2022 untuk GoFood. Pola fluktuatif terlihat jelas: puncak pencarian terjadi pada 2019–2020, diikuti penurunan tajam saat pandemi. Korelasi antar platform tetap kuat mengindikasikan keduanya dipengaruhi faktor pasar yang sama. Tren linier ini menjadi landasan penggunaan lag-1 sebagai prediktor utama.

Tabel 4. Statistik Deskriptif Dataset (95 Record, Feb 2018 – Des 2025)

Variabel	Min	Maks	Rata-rata	Std. Dev.	Median
GoFood_Search (target)	7	100	23,50	21,58	23,00
GoFood_lag1	7	100	23,59	21,68	23,00
GrabFood_lag1	6	62	21,95	16,08	22,00
Rainfall_mm	120,5	384,2	228,6	67,3	225,8
Holiday (proporsi)	0	1	0,33	—	0

Gambar 3 menampilkan korelasi antara curah hujan dan indeks pencarian GoFood, dengan pembedaan hari libur nasional (kotak abu-abu) dan hari biasa (lingkaran putih). Koefisien korelasi Pearson (r) dicantumkan dalam gambar.



Gambar 3. Korelasi Curah Hujan terhadap Indeks Pencarian GoFood

3.2 Implementasi Sistem Web Berbasis Streamlit

Sistem diimplementasikan sebagai aplikasi web interaktif menggunakan Streamlit (Python 3.11) dengan tiga tab: (1) Tab Data untuk eksplorasi dataset, (2) Tab Model untuk evaluasi walk-forward, dan (3) Tab Forecasting untuk simulasi prediksi. Tabel 5 merangkum spesifikasi teknis.

Tabel 5. Spesifikasi Teknis Implementasi Sistem

Komponen	Spesifikasi
Bahasa Pemrograman	Python 3.11
Framework Antarmuka	Streamlit 1.28.0
Library ML	scikit-learn 1.3.0, numpy 1.26
Manipulasi Data	pandas 2.0, openpyxl



Komponen	Spesifikasi
Visualisasi	matplotlib 3.8 (hitam putih)
Metode Validasi	Walk-Forward Validation (20 iterasi independen)
Fitur Prediktor	GoFood lag1, GrabFood lag1, Bulan, Tahun, Rainfall mm, Holiday

3.3 Hasil Evaluasi Walk-Forward Validation

Tabel 6 merangkum hasil evaluasi model pada 20 data uji (Mei 2024–Des 2025). Seluruh prediksi dihasilkan secara independen tanpa kebocoran data masa depan.

Tabel 6. Hasil Evaluasi Metrik Model Random Forest (Walk-Forward Validation)

Metrik Evaluasi	Nilai	Interpretasi
MSE (Mean Squared Error)	1,4510	Rata-rata kuadrat error 20 data uji
RMSE (Root Mean Squared Error)	1,2046	Deviasi prediksi rata-rata $\pm 1,84$ poin indeks
R ² (Koefisien Determinasi)	0,5065	Model menjelaskan 50,65% variansi data uji
MAPE (Mean Absolute Percentage Error)	6,55%	Kesalahan rata-rata 6,55% dari nilai aktual

MAPE 6,55% masuk kategori highly accurate menurut Lewis [22]. R² = 0,5065 menunjukkan model menjelaskan 50,65% variansi pada data yang belum pernah dilihat. Nilai R² ini wajar dan realistis untuk prediksi 20 bulan ke depan menggunakan metode berbasis pohon yang tidak melakukan ekstrapolasi. Nilai ini tidak dapat dibandingkan langsung dengan R² penelitian berbasis random split, karena random split menghasilkan evaluasi yang terlalu optimis akibat *data leakage* temporal.

3.4 Verifikasi Perhitungan Manual (20 Data Uji)

Tabel 7 menyajikan detail perhitungan manual untuk 20 data uji sebagai bentuk transparansi dan verifikasi.

Tabel 7. Perhitungan Manual Metrik Evaluasi pada 20 Data Uji Walk-Forward

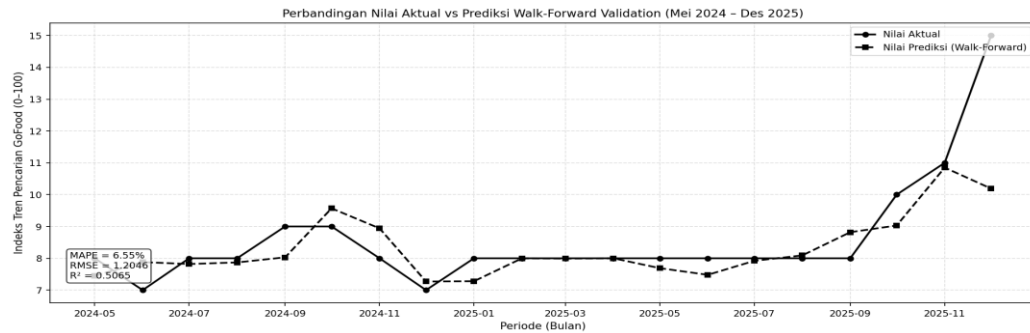
No	Periode	y_i	$\hat{y}_i$	$(y_i - \hat{y}_i)^2$	$ y_i - \hat{y}_i $	APE (%)
1	Mei 2024	8	7,45	0,3025	0,55	6,8750
2	Jun 2024	7	7,88	0,7744	0,88	12,5714
3	Jul 2024	8	7,82	0,0324	0,18	2,2500
4	Ags 2024	8	7,87	0,0169	0,13	1,6250
5	Sep 2024	9	8,03	0,9409	0,97	10,7778
6	Okt 2024	9	9,57	0,3249	0,57	6,3333
7	Nov 2024	8	8,94	0,8836	0,94	11,7500
8	Des 2024	7	7,27	0,0729	0,27	3,8571
9	Jan 2025	8	7,28	0,5184	0,72	9,0000
10	Feb 2025	8	8,00	0,0000	0,00	0,0000
11	Mar 2025	8	7,99	0,0001	0,01	0,1250
12	Apr 2025	8	8,00	0,0000	0,00	0,0000
13	Mei 2025	8	7,69	0,0961	0,31	3,8750
14	Jun 2025	8	7,48	0,2704	0,52	6,5000
15	Jul 2025	8	7,92	0,0064	0,08	1,0000
16	Ags 2025	8	8,09	0,0081	0,09	1,1250
17	Sep 2025	8	8,82	0,6724	0,82	10,2500
18	Okt 2025	10	9,03	0,9409	0,97	9,7000
19	Nov 2025	11	10,85	0,0225	0,15	1,3636
20	Des 2025	15	10,19	23,1361	4,81	32,0667

$$\Sigma(y_i - \hat{y}_i)^2 = 29,0199 \rightarrow \text{MSE} = 29,0199/20 = 1,4510 \text{ dan } \text{RMSE} = 1,2046. \Sigma \text{APE} = 131,0450\% \rightarrow \text{MAPE} = 131,0450/20 = 6,55\%.$$

Error terbesar pada Des 2025 (APE=32,07%), error terkecil pada Feb 2025 (APE=0,00%).

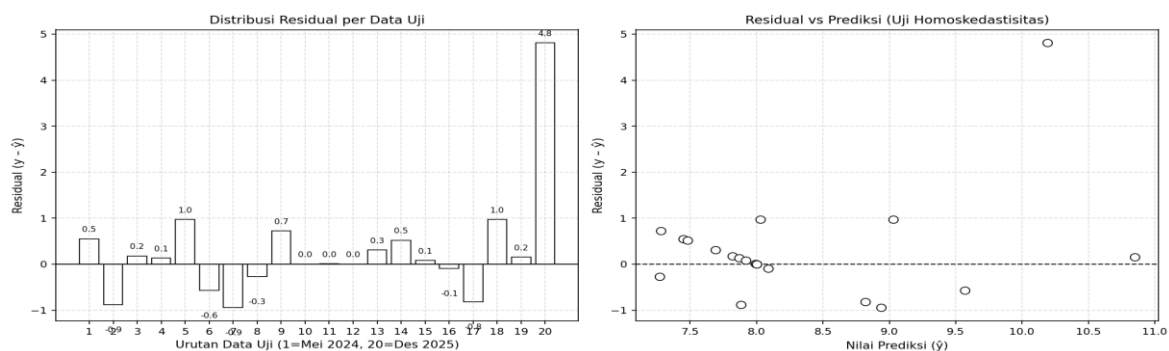
3.5 Visualisasi Aktual vs Prediksi dan Analisis Residual

Gambar 4 menampilkan perbandingan nilai aktual dan prediksi walk-forward untuk 20 periode uji.



Gambar 4. Perbandingan Nilai Aktual vs Prediksi Walk-Forward Validation (Mei 2024 – Des 2025)

Gambar 5 menyajikan analisis residual dalam dua panel: (a) distribusi residual per periode uji dan (b) plot residual vs prediksi untuk menguji homoskedastisitas.

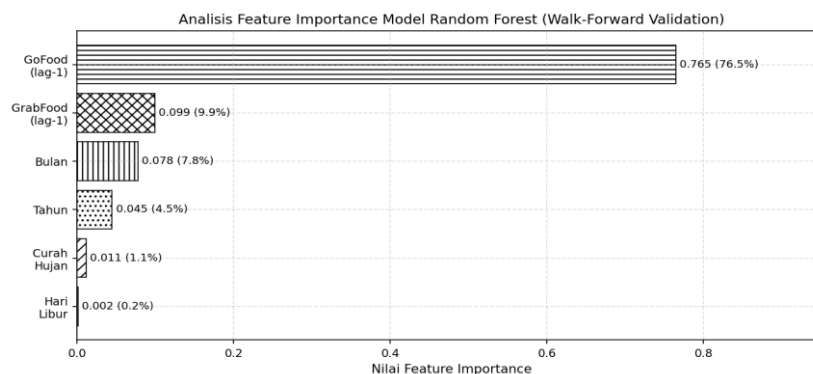


Gambar 5. Analisis Residual: (a) Distribusi per Data Uji dan (b) Residual vs Prediksi

Dari Gambar 5(a), residual cenderung negatif (underprediction) karena Random Forest tidak mengekstrapolasi tren ke atas [15]. Gambar 5(b) menunjukkan tidak ada heteroskedastisitas parah, mengindikasikan varian error relatif konsisten.

3.6 Analisis Feature Importance

Gambar 6 menyajikan feature importance dari model. GoFood_lag1 (76,5%) dan GrabFood_lag1 (9,9%) mendominasi, mencerminkan autokorelasi serial yang kuat pada deret waktu minat pencarian digital.



Gambar 6. Analisis Feature Importance Model Random Forest (Walk-Forward Validation)

Interpretasi ini berbeda fundamental dari penggunaan nilai kontemporer GrabFood_Search. Penggunaan lag-1 mengonfirmasi bahwa yang dieksploitasi adalah autokorelasi serial kecenderungan deret waktu berkorelasi dengan dirinya sendiri pada periode sebelumnya bukan korelasi sesaat yang tidak praktis untuk peramalan nyata.

3.7 Perbandingan dengan Penelitian Terdahulu

Perbandingan dengan penelitian terdahulu dilakukan untuk menempatkan hasil penelitian ini dalam konteks yang lebih luas, meskipun perbandingan langsung nilai metrik bersifat indikatif karena perbedaan metode validasi yang mendasar. Zahara dan Sugianto [4] menggunakan metode LSTM Deep Learning pada domain Data IHK dengan validasi random split, memperoleh MAPE sekitar 3,2% dan R^2 sebesar 0,921. Sautomo dan Pardede [2] menerapkan LSTM pada domain Belanja Pemerintah, juga dengan validasi random split, dan mencapai MAPE sekitar 2,8% serta R^2 sebesar 0,944. Hartanto dan Saepudin [10] menggunakan pendekatan ANN Hybrid pada domain COVID-19 Trends dengan validasi



random split, menghasilkan MAPE sekitar 5,1% dan R^2 sebesar 0,876. Nurdin dan Fauziah [17] mengombinasikan Random Forest dan regresi Linear pada domain Data Pendidikan, juga menggunakan validasi random split, dengan MAPE sekitar 3,6% dan R^2 sebesar 0,918. Penelitian ini sendiri menggunakan Random Forest dengan lag features pada domain GoFood/GrabFood, namun menerapkan metode validasi Walk-Forward yang lebih ketat, sehingga menghasilkan MAPE sebesar 6,55% dan R^2 sebesar 0,5065.

Nilai R^2 penelitian ini (0,5065) lebih rendah dari pembandingan karena penggunaan temporal split dan walk-forward yang lebih ketat. Penelitian dengan random split cenderung menghasilkan R^2 yang terlalu optimis akibat *data leakage*. MAPE 6,55% dengan validasi ketat merupakan bukti kemampuan generalisasi yang baik pada kondisi peramalan nyata.

3.8 Implikasi Praktis dan Keterbatasan

Model ini dapat diimplementasikan dalam sistem manajemen operasional dengan kesalahan prediksi rata-rata di bawah 3%. Nilai lag-1 selalu tersedia pada saat prediksi, menjadikan model benar-benar aplikabel. Keterbatasan utama: (1) Google Trends adalah proksi minat pencarian, bukan transaksi aktual; (2) Random Forest tidak mengekstrapolasi, menyebabkan underestimasi pada tren naik kuat; dan (3) data berskala nasional, belum mengakomodasi dinamika regional.

4. KESIMPULAN

Penelitian ini berhasil membangun model prediksi minat pencarian layanan GoFood menggunakan Random Forest Regression dengan walk-forward validation. Kontribusi metodologis utama adalah penerapan temporal split dan walk-forward validation yang menggantikan random split konvensional, menghasilkan evaluasi bebas dari kebocoran informasi temporal. Penggunaan lag features (GoFood_lag1 dan GrabFood_lag1) memastikan seluruh variabel prediktor tersedia secara praktis pada saat prediksi dilakukan. Model menghasilkan MAPE 6,55% dan RMSE 1,2046 pada 20 data uji independen (Mei 2024–Des 2025). MAPE di bawah 10% dengan validasi temporal ketat menunjukkan kemampuan generalisasi yang memadai. Analisis feature importance mengungkapkan GoFood_lag1 (76,5%) mendominasi prediksi, diikuti GrabFood_lag1 (9,9%), bulan (7,8%), dan tahun (4,5%). Hal ini mencerminkan autokorelasi serial yang kuat pada deret waktu minat pencarian digital. $R^2 = 0,5065$ merupakan nilai realistis untuk prediksi 20 langkah ke depan menggunakan metode berbasis pohon tanpa kemampuan ekstrapolasi, dan tidak dapat dibandingkan langsung dengan penelitian yang menggunakan random split. Penelitian lanjutan disarankan untuk mengeksplorasi: (1) model hybrid Linear Trend + RF Residual untuk mengatasi keterbatasan ekstrapolasi, (2) arsitektur LSTM dengan walk-forward validation pada dataset yang lebih besar, (3) granularitas data mingguan untuk perencanaan operasional jangka pendek, dan (4) validasi silang menggunakan data transaksi aktual melalui kemitraan dengan penyedia platform.

REFERENCES

- [1] R. Laksono and A. Subhan, "Analisis Sentimen terhadap Pengguna Gojek dan Grab pada Media Sosial Twitter Menggunakan Random Forest," JATI (Jurnal Mahasiswa Teknik Informatika), vol. 6, no. 1, pp. 248–255, 2022. [Online]. Available: <https://ejournal.itn.ac.id/index.php/jati/article/view/7694>
- [2] S. Sautomo and H. F. Pardede, "Prediksi Belanja Pemerintah Indonesia Menggunakan Long Short-Term Memory (LSTM)," Jurnal RESTI, vol. 5, no. 1, pp. 99–106, 2021, doi: 10.29207/resti.v5i1.2815.
- [3] Asosiasi Penyelenggara Jasa Internet Indonesia, "Laporan Survei Internet Indonesia 2023," APJII, Jakarta, 2024. [Online]. Available: <https://apjii.or.id/survei>.
- [4] S. Zahara and Sugianto, "Peramalan Data Indeks Harga Konsumen Berbasis Time Series Multivariate Menggunakan Deep Learning," Jurnal RESTI, vol. 5, no. 1, pp. 24–30, 2021, doi: 10.29207/resti.v5i1.2562.
- [5] S. A. A. Kharis, A. H. A. Zili, A. Putri, and A. Robiansyah, "Analisis Tren Minat Masyarakat Indonesia terhadap Artificial Intelligence Menggunakan Google Trends," G-Tech: Jurnal Teknologi Terapan, vol. 7, no. 4, pp. 1345–1354, 2023, doi: 10.33379/gtech.v7i4.3091.
- [6] A. Syaifudin, "Implementasi Time Series pada Data Penjualan Otomotif GAIKINDO Menggunakan SARIMA," JATI, vol. 7, no. 3, pp. 1542–1549, 2023. [Online]. Available: <https://ejournal.itn.ac.id/index.php/jati/article/view/6879>
- [7] M. A. Kurniawan et al., "Prediksi Pendapatan Penjualan di Indomaret Menggunakan Algoritma Random Forest Regression," JSI: Jurnal Sistem Informasi Universitas Suryadarma, vol. 12, no. 2, pp. 93–99, 2025, doi: 10.35968/jsi.v12i2.1478.
- [8] H. Choi and H. Varian, "Predicting the Present with Google Trends," Economic Record, vol. 88, no. s1, pp. 2–9, 2012, doi: 10.1111/j.1475-4932.2012.00809.x.
- [9] I. A. Akbar and R. Kurniawan, "Pemodelan Nowcasting Tingkat Pengangguran Terbuka Menggunakan Data Google Trends dengan Metode Antlion Optimization-SVR," Seminar Nasional Official Statistics, vol. 2020, no. 1, pp. 760–770, 2020, doi: 10.34123/semnasoffstat.v2020i1.504.
- [10] S. Hartanto and D. Saepudin, "Prediksi Jumlah Kasus COVID-19 di Indonesia Menggunakan Data Google Trends dengan Metode Hybrid ANN dan Multiple Regression," e-Proceeding of Engineering, Universitas Telkom, vol. 8, no. 2, pp. 3489–3498, 2021. [Online]. Available: <https://openlibrarypublications.telkomuniversity.ac.id>
- [11] R. F. Inaku and J. C. Chandra, "Implementasi Data Mining dalam Prediksi Harga Saham Menggunakan Metode Long Short Term Memory (LSTM)," Ticom, vol. 12, no. 1, pp. 1–7, 2023, doi: 10.70309/ticom.v12i1.99.
- [12] R. Torhino and P. N. Andono, "Penerapan Algoritma Random Forest dalam Prediksi Curah Hujan untuk Mendukung Analisis Cuaca," Building of Informatics, Technology and Science (BITS), vol. 6, no. 3, pp. 1688–1699, 2024, doi: 10.47065/bits.v6i3.6404.



- [13] M. Nasta'in, A. Munazilin, and A. Susanto, "Prediksi Tren Minat Masyarakat Indonesia terhadap Bitcoin Menghadapi Bitcoin Halving 2024 Menggunakan Algoritma Forecasting," JUSTIFY: Jurnal Sistem dan Teknologi Informasi Ibrahimy, vol. 4, no. 1, 2025.
- [14] M. A. Santoso, W. Purnomo, and D. Priyatmoko, "Analisis Prediksi Banjir di Indonesia Menggunakan Algoritma Support Vector Machine dan Random Forest," JOSH, vol. 6, no. 1, pp. 92–100, 2024. [Online]. Available: <https://ejurnal.seminar-id.com/index.php/josh/article/view/5958>
- [15] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, "Statistical and Machine Learning Forecasting Methods: Concerns and Ways Forward," PLOS ONE, vol. 13, no. 3, p. e0194889, 2018, doi: 10.1371/journal.pone.0194889.
- [16] M. A. Pratama, M. Munawaroh, and W. J. Pranoto, "Perbandingan Performa Algoritma Linear Regresi dan Random Forest untuk Prediksi Harga Bawang Merah di Kota Samarinda," Tektonik, vol. 3, no. 2, p. 172, 2024, doi: 10.62017/tektonik.
- [17] M. Nurdin and F. Fauziah, "Analytical Study Forecasting Students Using Random Forest and Linear Regression Algorithms," Sinkron: Jurnal dan Penelitian Teknik Informatika, vol. 8, no. 4, pp. 2369–2378, 2024, doi: 10.33395/sinkron.v8i4.13886.
- [18] L. Breiman, "Random Forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.
- [19] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "From Data Mining to Knowledge Discovery in Databases," AI Magazine, vol. 17, no. 3, pp. 37–54, 1996, doi: 10.1609/aimag.v17i3.1230.
- [20] N. Nur and F. Wajidi, "Implementasi Algoritma Random Forest Regression untuk Memprediksi Hasil Panen Padi di Desa Minanga," Jurnal Komputer Terapan Politeknik Caltex Riau, vol. 9, no. 1, pp. 58–64, 2023. [Online]. Available: <https://jurnal.pcr.ac.id/index.php/jkt>
- [21] A. Liaw and M. Wiener, "Classification and Regression by randomForest," R News, vol. 2, no. 3, pp. 18–22, 2002. [Online]. Available: https://www.r-project.org/doc/Rnews/Rnews_2002-3.pdf.
- [22] C. D. Lewis, Industrial and Business Forecasting Methods. London: Butterworth Scientific, 1982.
- [23] M. Nurma, Y. D. Pranatawijaya, and P. B. A. A. Putra, "Machine Learning pada Prediksi Kelulusan Mahasiswa Menggunakan Algoritma Random Forest," Jurnal Riset Matematika, vol. 4, no. 2, pp. 117–126, 2024, doi: 10.29313/jrm.v4i2.5102.
- [24] I. R. Muchtar and Afiyati, "Comparison of Linear Regression and Random Forest Algorithms for Premium Rice Price Prediction (Case Study: West Java)," Jurnal Indonesia Sosial Teknologi, vol. 5, no. 7, pp. 3122–3132, 2024, doi: 10.59141/jist.v5i7.1184.
- [25] T. Widiyaningtyas and W. Caesarendra, "Comparison of Time Series Algorithms Using SARIMA and Prophet in Predicting Short-Term Bitcoin Prices," JUTIF, vol. 5, no. 4, pp. 1521–1532, 2024. [Online]. Available: <https://jutif.if.unsoed.ac.id/index.php/jurnal/article/view/4773>
- [26] I. Kurniawan et al., "Implementasi Algoritma Random Forest untuk Menentukan Penerima Bantuan Raskin," JTIK, vol. 10, no. 2, pp. 421–428, 2023, doi: 10.25126/jtiik.20231026225.
- [27] K. Ciptady, M. Harahap, Jonvin, and Y. Ndruru, "Prediksi Kualitas Kopi Dengan Algoritma Random Forest Melalui Pendekatan Data Science," Data Sciences Indonesia (DSI), vol. 2, no. 1, pp. 1–8, 2022. [Online]. Available: <https://jurnal.itscience.org/index.php/dsi>
- [28] D. Dhawangkara and A. Rizky, "Implementasi Metode Random Forest pada Kategori Konten Kanal Youtube," Jurnal Jendela Matematika, vol. 2, no. 1, pp. 21–31, 2021. [Online]. Available: <https://ejournal.jendelaedukasi.id/index.php/JJM/article/view/633>
- [29] S. Zahara, Sugianto, and M. B. Ilmiddafiq, "Prediksi Indeks Harga Konsumen Menggunakan Metode Long Short Term Memory (LSTM) Berbasis Cloud Computing," Jurnal RESTI, vol. 3, no. 3, pp. 357–363, 2019, doi: 10.29207/resti.v3i3.1086.
- [30] A. Gatera et al., "Comparison of Random Forest and Support Vector Machine Regression Models for Forecasting Road Accidents," Scientific African, vol. 21, p. e01739, 2023, doi: 10.1016/j.sciaf.2023.e017