Terbit online pada laman web jurnal: **https://journal.upp.ac.id/index.php/rjti/index**

RJTI
Riau Jurnal Teknik Informatika

Vol. 4 No. 3 (2025) 446 – 453 | ISSN: 2829-2529 (Media Online)

Implementasi Algoritma Random Forest untuk Prediksi Volume Pengunjung pada Sektor Layanan Publik (Studi Kasus Unit Layanan Publik di Jawa Barat Tahun 2021-2022)

Anisa Pebriyani Huslan[1], Fathoni Mahardika[2], Dani Indra Junaedi[3]
[1,2,3]Informatika, Fakultas Teknologi Informasi, Universitas Sebelas April
[1]220660121016@student.unsap.ac.id, [2]fathoni@unsap.ac.id, [3]dani@unsap.ac.id

Abstract

Accurately anticipating daily visitor volumes is pivotal for reliable staffing, short queues, and better citizen experience in public services. This study investigates a practical and reproducible approach that relies solely on calendar signals to forecast daily arrivals using a Random Forest. We analyze a daily dataset for public service units from 2021 to 2022 with 730 observations, split chronologically into 80 percent training and 20 percent testing to mirror real deployment and avoid leakage. Predictors include day, month, day of week, weekend indicator, week in month, and month start and end flags. The model employs 200 trees with max_depth = 8. Performance is evaluated with MAE, MAPE, RMSE, and R^2, and also compared against two strong baselines, namely naive lag-1 and day of week mean. On the test set, the model attains $R^2 = 0.873$, RMSE = 16.328, MAE = 13.103, and MAPE about 5.00 percent, outperforming both baselines. Quantitatively, the model reduced prediction errors (MAPE) by about 64% and 67% compared to the naïve lag-1 and day-of-week baselines, respectively, highlighting its empirical superiority. Feature importance indicates day of week and weekend as the most influential variables, aligning with operational intuition. The results demonstrate that a low cost, data light forecast can deliver decision grade accuracy that is ready for front line scheduling, queue risk monitoring, and capacity allocation, while offering a transparent path for scaling with additional context such as holidays and weather.

Keywords: Random Forest, Time Series Forecasting, Calendar Features, Public Service, Visitor Volume

Abstrak

Peramalan volume pengunjung harian yang akurat merupakan kunci untuk penjadwalan petugas yang andal, antrean yang singkat, dan pengalaman warga yang lebih baik pada layanan publik. Penelitian ini mengajukan pendekatan praktis dan mudah direplikasi yang hanya memanfaatkan sinyal kalender untuk meramalkan kedatangan harian menggunakan *Random Forest*. Dataset harian unit pelayanan publik tahun 2021 hingga 2022 berisi 730 observasi dan dibagi secara kronologis menjadi 80 persen pelatihan dan 20 persen pengujian agar menyerupai penerapan nyata serta menghindari kebocoran informasi. Prediktor mencakup hari, bulan, hari dalam minggu, indikator akhir pekan, minggu dalam bulan, serta penanda awal dan akhir bulan. Model menggunakan 200 pohon dengan *max_depth* = 8. Kinerja dievaluasi menggunakan *MAE, MAPE, RMSE*, dan R^2 serta dibandingkan dengan dua *baseline* yang kuat, yaitu *naive lag*-1 dan rata-rata hari dalam minggu. Pada data uji, model mencapai $R^2 = 0.873$, RMSE = 16.328, MAE = 13.103, dan MAPE sekitar 5,00 persen, melampaui kedua *baseline*. Secara kuantitatif, model ini menurunkan kesalahan prediksi (MAPE) masing-masing sekitar 64% dan 67% dibandingkan *baseline naïve* lag-1 dan rata-rata hari dalam minggu, menegaskan keunggulan empirisnya. Analisis *feature importance* menempatkan hari dalam minggu dan akhir pekan sebagai variabel paling berpengaruh dan sejalan dengan intuisi operasional. Temuan ini menunjukkan bahwa peramalan berbiaya rendah dan hemat data mampu memberikan akurasi yang siap pakai untuk penjadwalan, pemantauan risiko antrean, dan alokasi kapasitas, serta mudah ditingkatkan dengan konteks tambahan seperti hari libur dan cuaca.

Kata kunci: *Random Forest, Peramalan Deret Waktu, Fitur Kalender, Layanan Publik, Volume Pengunjung*

1. Pendahuluan

Pelayanan publik modern dituntut menjaga mutu pengalaman warga melalui penjadwalan petugas dan pengaturan kapasitas yang presisi di tengah permintaan harian yang fluktuatif. Literatur administrasi publik dan tata kelola digital menunjukkan bahwa analitik prediktif dan *artificial intelligence* (AI) berperan meningkatkan kapasitas antisipatif pemerintah, termasuk peramalan kedatangan harian dan pengelolaan beban unit garis depan [1]–[7]. Sejalan dengan itu, laporan kebijakan

internasional menegaskan korelasi antara kematangan tata kelola digital dan kemampuan proaktif dalam perencanaan dan penjadwalan layanan [1]–[3]. Kondisi tersebut menegaskan kebutuhan akan pendekatan peramalan yang akurat, mudah dioperasionalkan, dan tidak bergantung pada prasyarat data kompleks agar dapat diadopsi lintas instansi.

Tantangan utama pada unit layanan publik berkaitan dengan fluktuasi permintaan yang dipengaruhi faktor kalender hari kerja dan hari libur, siklus musiman, kebijakan operasional, serta dinamika mobilitas. Hal ini terkonfirmasi pada unit layanan publik yang menjadi studi kasus penelitian. Data operasional harian (2021-2022) menunjukkan variabilitas yang ekstrem, di mana tingkat kunjungan harian dapat berkisar dari 80 pengunjung pada hari sepi hingga melonjak melampaui 500 pengunjung pada hari puncak layanan. Tanpa peramalan yang memadai, lonjakan atau penurunan tak terantisipasi berpotensi menimbulkan antrean panjang, ketidakseimbangan beban kerja, dan inefisiensi biaya. Literatur mengenai antrean menegaskan bahwa prediksi kedatangan yang andal dapat diintegrasikan dengan kebijakan penjadwalan dan perencanaan kapasitas untuk menurunkan waktu tunggu dan memaksimalkan penggunaan kapasitas layanan [8].

Pemodelan berbasis *machine learning* (ML) telah diterapkan luas pada domain layanan publik dengan hasil yang menjanjikan di transportasi, kesehatan, dan pengelolaan ruang publik [9]–[11]. Pada transportasi publik, kajian sistematis dan studi aplikatif menunjukkan bahwa model ML efektif memodelkan serta meramalkan permintaan penumpang dalam kondisi normal maupun disrupsi [9], [12]–[14]. Pada layanan kesehatan, penelitian melaporkan peningkatan akurasi untuk prakiraan kedatangan instalasi gawat darurat, prediksi admisi, prediksi volume rawat jalan, dan prediksi ketidakhadiran janji temu yang berdampak pada penjadwalan dan efisiensi operasional [15]–[19]. Pada ruang publik dan destinasi, pemodelan kunjungan pada skala harian hingga beresolusi tinggi membantu strategi kapasitas yang adaptif serta mitigasi kepadatan [20]–[22]. Namun, banyak studi mengandalkan data eksogen yang kaya atau arsitektur kompleks sehingga menimbulkan hambatan implementasi. Kesenjangan ini adalah bukti empiris tentang kelayakan model yang hemat data dan berbiaya rendah, khususnya pendekatan kalender saja yang dievaluasi secara operasional pada data kecil khas administrasi publik.

Secara metodologis, model *Random Forest* (RF) merupakan kandidat kuat untuk data tabular berukuran kecil hingga menengah yang lazim dijumpai pada administrasi publik. *Benchmark* pada ranah tabular menunjukkan bahwa model berbasis pohon sering kali unggul dibandingkan alternatifnya, terutama ketika jumlah fitur terbatas, relasi bersifat non-linier, dan terdapat interaksi antar fitur [23], [24]. Kerangka *explainable AI* memungkinkan penelusuran kontribusi fitur sehingga keluaran model dapat diinterpretasikan oleh pemangku kepentingan [25], [26]. Dalam konteks deret waktu, perluasan menuju regresi kuantil juga membuka ruang estimasi ketidakpastian bagi keputusan berisiko [26]. Prinsip peramalan dan rekayasa fitur menyediakan acuan untuk pemilihan metrik, pemisahan berbasis waktu, dan konstruksi fitur yang efisien [27]–[33].

Algoritma *Random Forest* ini telah menunjukkan efektivitas yang signifikan dalam berbagai domain, khususnya pada sektor komersial dan bisnis. Sejumlah penelitian telah berhasil menerapkannya untuk memprediksi permintaan dalam sektor komersial, seperti restoran, coffee shop, dan bisnis retail [11], [34], [35]. Kemampuan prediktifnya mampu memberikan *insight* berguna untuk mengidentifikasi produk atau layanan yang akan paling diminati, sehingga memungkinkan optimalisasi manajemen inventori dan perencanaan ketersediaan stok [35], [36]. *Random Forest* banyak dipilih dalam tugas prediksi karena kemampuannya menangani data non-linear, mengurangi *overfitting*, serta menghasilkan prediksi yang stabil [11], [35], [36]. Studi serupa juga menegaskan bahwa *Random Forest* memiliki tingkat akurasi yang tinggi dengan kesalahan prediksi yang relatif rendah serta interpretasi yang mudah [10]. Dengan mempertimbangkan efektivitas algoritma *Random Forest* dalam memodelkan pola permintaan pada konteks komersial, penerapan pendekatan serupa pada sektor publik berpotensi memberikan kontribusi signifikan dalam mendukung pengambilan keputusan berbasis data terhadap peningkatan efisiensi operasional dan kualitas layanan kepada masyarakat secara proaktif [7], [31].

Penelitian ini menyajikan *proof of concept* penerapan *Random Forest* berbasis fitur waktu untuk peramalan volume pengunjung harian pada sebuah unit layanan publik. Desain fitur berangkat dari informasi yang lazim tersedia dan rendah hambatan implementasi, yaitu indikator kalender (tahun, bulan, hari, hari dalam minggu, penanda akhir pekan, dan hari ke dalam tahun) serta sinyal historis yang dihitung ketat dari data masa lalu, meliputi *lag*-1, *lag*-7, *rolling mean* tujuh hari (*MA7*), dan *rolling mean* empat belas hari (*MA14*). Dataset terdiri atas deret waktu harian tahun 2021 sampai 2022 dengan jumlah observasi sekitar $n \approx 730$. Evaluasi menggunakan pemisahan berbasis waktu dengan blok terakhir sebagai *holdout* uji dan pembandingan terhadap dua *baseline* operasional yang banyak digunakan, yaitu *naïve lag*-1 dan rata-rata hari dalam minggu [27], [30], [32], [33].

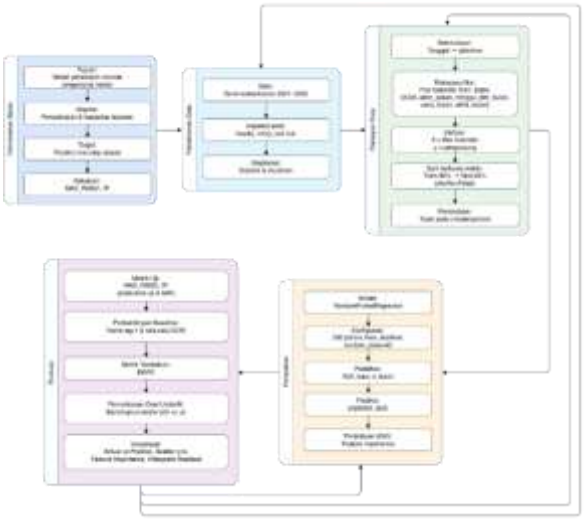
Tujuan penelitian ini adalah untuk (1) mengembangkan serta mengevaluasi model peramalan volume pengunjung harian berbasis *Random Forest* dengan

DOI : https://doi.org/10.52158/rjti.v4i3.4042

memanfaatkan fitur kalender sederhana pada data berukuran kecil, (2) membandingkan kinerjanya dengan *baseline* operasional yang umum digunakan (*naïve lag*-1 dan rata-rata hari dalam minggu) agar manfaatnya terukur, serta (3) menyajikan analisis tingkat fitur yang memungkinkan interpretasi oleh pemangku kepentingan sehingga keluaran model dapat ditranslasikan ke dalam kebijakan penjadwalan dan pengaturan kapasitas.

2. Metode Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan kerangka *CRISP-DM* (*Cross-Industry Standard Process for Data Mining*) hingga fase *Evaluation*, sedangkan fase *Deployment* dibahas sebagai implikasi. Pemilihan kerangka ini selaras dengan tata kelola layanan publik berbasis data dan AI yang menekankan akuntabilitas proses dari tujuan kebijakan hingga validasi model [1]–[7]. Tahapan pada penelitian ditunjukkan pada Gambar 1.



Gambar 1. Tahapan Penelitian (CRISP-DM)

Pemahaman Bisnis (*Business Understanding*)

Tujuan bisnis ini adalah menyediakan model peramalan volume pengunjung harian yang akurat, stabil, dan mudah untuk dioperasionalkan dalam mendukung penjadwalan petugas dan pengaturan kapasitas. Relevansi kebutuhan ini tercermin pada literatur kebijakan digital dan manajemen antrean/operasional [1]–[3], [8], [9], [13], [14]. Keluaran yang dituju yaitu berupa deret prediksi dengan ukuran akurasi MAE, MAPE, RMSE, dan R^2 yang mudah ditafsirkan oleh pemangku kepentingan.

Pemahaman Data (*Data Understanding*)

Dataset berupa data deret waktu jumlah pengunjung harian tahun 2021–2022 pada salah satu unit pelayanan publik di Jawa Barat dengan variabel inti tanggal dan jumlah pengunjung (n = 730). Karakterisasi awal meliputi: pratinjau struktur data (*df.head*, *df.info*), pemeriksaan nilai hilang, deskriptif ringkas (min, p25, median, mean, p75, max), dan inspeksi musiman mingguan/kalender. Untuk konteks reproduksibilitas, unit observasi adalah hari kalender pada satu unit layanan publik yang sama. Praktik ini mengikuti anjuran literatur *open data* dan kompetisi peramalan mengenai transparansi struktur data, pencegahan *leakage*, serta pemisahan berbasis waktu [27], [29]–[33].

Persiapan Data (*Data Preparation*)

Persiapan data diawali dengan normalisasi tipe tanggal, yaitu mengonversi kolom tanggal ke tipe *datetime* agar dapat digunakan untuk merekayasa fitur berbasis kalender. Selanjutnya dibentuk sekumpulan fitur berbiaya rendah yang lazim tersedia dan tidak menimbulkan *data leakage*, mencakup hari (1–31), bulan (1–12), hari_dlm_minggu (1–7), akhir_pekan (indikator Sabtu/Minggu), minggu_dlm_bulan (1–5), awal_bulan (hari ke-1 sampai ke-7), serta akhir_bulan (tujuh hari terakhir di suatu bulan). Rancangan fitur ini bertujuan menangkap pola musiman mingguan dan kalender yang dilaporkan dominan pada berbagai layanan publik dan destinasi, sekaligus konsisten dengan praktik rekayasa fitur deret waktu dan otomasi konstruksi fitur pada literatur mutakhir [9], [11]–[13], [15], [20]–[22], [28], [32], [33].

Seluruh fitur kalender tersebut dihimpun sebagai prediktor X, sedangkan variabel target y adalah jumlah pengunjung. Untuk meniru skenario prediksi ke depan dan mencegah *look-ahead bias*, pemisahan data dilakukan secara berbasis waktu menggunakan *train–test split* 80/20 yang dipertahankan dalam urutan kronologis (*time-based split*, *shuffle* = *False*), sesuai anjuran kompetisi peramalan dan pedoman evaluasi empiris [29], [30]. Mengingat model yang digunakan berbasis pohon, penskalaan tidak dilakukan karena algoritma tersebut relatif tidak sensitif terhadap perbedaan skala dan lazim unggul pada data tabular [23], [24].

Pemodelan (*Modeling*)

Pemodelan pada penelitian ini dilakukan dengan *RandomForestRegressor* yang dikonfigurasi dengan menggunakan 200 pohon (n_estimators = 200), 8 untuk kedalaman maksimum (max_depth = 8), serta 42 untuk *random seed* (random_state = 42). Konfigurasi *hyperparameters* model, khususnya n_estimators = 200 dan max_depth = 8, diperoleh melalui proses *tuning* yang sistematis. Proses *tuning* ini menerapkan metode *Grid Search* yang dikombinasikan dengan *Time-Series Cross-Validation* (Validasi Silang Deret Waktu) pada himpunan data pelatihan (data 80% awal) untuk menemukan kombinasi yang meminimalkan *Root Mean Squared Error* (RMSE). Adapun random_state = 42

DOI : https://doi.org/10.52158/rjti.v4i3.4042

ditetapkan untuk memastikan keterulangan (reprodusibilitas) hasil.

Pemilihan *Random Forest* didukung temuan komparatif yang menunjukkan keunggulan keluarga model berbasis pohon pada data tabular, termasuk ketahanannya terhadap nonlinieritas dan interaksi antar fitur [23], [24]. Model dilatih pada himpunan pelatihan melalui model.fit (X_train, y_train) dan menghasilkan prediksi pada himpunan uji melalui model.predict (X_test). Untuk mendukung keterjelasan hasil bagi pengambil kebijakan, digunakan *feature importance* guna menelaah kontribusi relatif setiap fitur terhadap keputusan model [25], [34].

Evaluasi (*Evaluation*)

Evaluasi kinerja dilakukan pada *holdout* uji menggunakan empat metrik utama, yaitu *Mean Absolute Error* (MAE), *Mean Absolute Percentage Error* (MAPE), *Root Mean Squared Error* (RMSE), dan koefisien determinasi (R^2). Metrik pada himpunan pelatihan turut dilaporkan untuk menilai indikasi *overfitting* atau *underfitting*, sejalan dengan konvensi evaluasi pada kompetisi peramalan dan studi empiris terkini [29], [30]. Untuk mengontekstualisasikan capaian model, disusun dua *baseline* operasional yang umum dipakai pada kajian deret waktu dan studi transportasi maupun pariwisata, yaitu *naïve lag-1* (nilai hari ini diasumsikan sama dengan nilai aktual hari sebelumnya) dan rata-rata *day-of-week* (DOW). Pada perbandingan ini juga dihitung *Mean Absolute Percentage Error* (*MAPE*) agar besaran galat mudah ditafsirkan oleh pemangku kepentingan [9], [13], [20], [22], [29].

Selain pelaporan numerik, empat visualisasi disiapkan untuk memperkaya penjelasan hasil, mencakup kurva aktual versus prediksi (periode uji), *scatter* prediksi terhadap aktual dengan garis identitas $y = x$, *feature importance* dari *Random Forest*, serta histogram *residuals*. Rangkaian visual ini membantu mendeteksi potensi bias sistematis, mengidentifikasi kasus ekstrem, dan mengaitkan temuan dengan kebutuhan operasional layanan publik [8], [25]. Seluruh proses dirancang replikabel dengan penetapan *random seed* (random_state = 42), pembagian data berbasis waktu 80/20 tanpa *shuffle* yang dinyatakan eksplisit, serta pelaporan MAE, MAPE, RMSE, dan R^2 pada pelatihan dan *holdout* uji. Pemeriksaan tipe data serta kelengkapan dilakukan sebelum pemodelan, sedangkan *feature importance* dan inspeksi *residuals* digunakan sebagai uji kewajaran hasil [25], [29], [30].

3. Hasil dan Pembahasan

Bagian ini memaparkan hasil pengujian kuantitatif dan pembahasan kualitatif terhadap model *Random Forest* berbasis fitur waktu kalender untuk peramalan volume pengunjung harian. Ringkasan jumlah observasi dan pembagian *temporal holdout* diuraikan terlebih dahulu. Selanjutnya disajikan kinerja model pada data pelatihan dan data pengujian, perbandingan terhadap *baseline* operasional, serta visualisasi kurva aktual versus prediksi, sebaran prediksi terhadap aktual, *feature importance*, dan distribusi *residuals*.

Ringkasan Data dan Pembagian *Holdout*

Pada Tabel 1 disajikan ringkasan jumlah observasi yang digunakan pada analisis. Total data harian yang siap dimodelkan berjumlah 730 observasi, dengan 584 observasi sebagai data pelatihan dan 146 observasi sebagai *holdout* uji. Skema pembagian ini mempertahankan keterurutan waktu (*shuffle = False*) agar evaluasi merepresentasikan penggunaan model pada konteks *future prediction*.

Tabel 1. Ringkasan Dataset dan Pembagian Data

Item	Nilai
Total Observasi	730
Jumlah Pelatihan (≈80%)	584
Jumlah Uji (≈20%)	146

Kinerja Model pada Data

Ringkasan metrik pada data pelatihan dan pengujian ditunjukkan pada Tabel 2. Pada *holdout* uji, model *Random Forest* mencapai $R^2 = 0.873$, RMSE = 16.328, MAE = 13.103, dan MAPE = 4.9996. Sedangkan, nilai R^2 pada data pelatihan sebesar 0.963 disertai RMSE = 10.101, MAE = 7.958, dan MAPE = 2.6309. Perbedaan kinerja pada data pelatihan dan pengujian berada pada rentang yang wajar, sehingga tidak mengindikasikan *overfitting* yang berat. Secara operasional, kesalahan absolut rata-rata sekitar 13 pengunjung per hari menunjukkan tingkat ketelitian yang memadai untuk mendukung penjadwalan petugas dan pengaturan kapasitas layanan.

Tabel 2. Kinerja Model *RF* pada Data Training dan Testing

Metrik	*Training*	*Testing*
R^2	0.963	0.873
RMSE	10.101	16.328
MAE	7.958	13.103
MAPE	2.6309	4.9996

Kinerja Model terhadap *Baseline* Operasional

Perbandingan kuantitatif kinerja antara *Random Forest* dan dua *baseline* operasional, yaitu *naïve lag*-1 dan rata-rata *day-of-week* (DOW), disajikan pada Tabel 3. Pada *holdout* uji, *Random Forest* memiliki MAPE 5,000% (≈ 4.9996%) dan $R^2 = 0.8733$, lebih baik daripada *naïve lag-1* (MAPE 13.8057%, $R^2 = -0.0815$) dan rata-rata DOW (MAPE 15,2756%, $R^2 = 0.0841$). Nilai MAE dan RMSE *Random Forest* juga paling rendah, yang menandakan penurunan kesalahan absolut dan kuadrat secara konsisten. Secara operasional, hasil ini bermakna

DOI : https://doi.org/10.52158/rjti.v4i3.4042

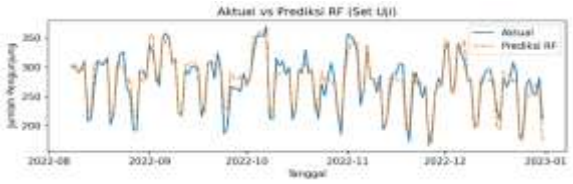
bahwa *Random Forest* berbasis fitur kalender mampu menurunkan kesalahan relatif lebih dari separuh dibandingkan dengan pendekatan praktis yang umum digunakan (*baseline*), sehingga berpotensi dalam meningkatkan ketepatan penjadwalan petugas dan pengaturan kapasitas layanan.

Tabel 3. Kinerja Model pada Set Uji (Perbandingan *Baseline*)

Model	MAE	RMSE	MAPE (%)	R^2
Random Forest	13.1031	16.3285	4.9996	0.8733
Naïve lag-1	34.8562	47.714	13.8057	-0.0815
Rata-rata DOW	38.8357	43.909	15.2756	0.0841

Analisis Kurva Aktual vs Prediksi

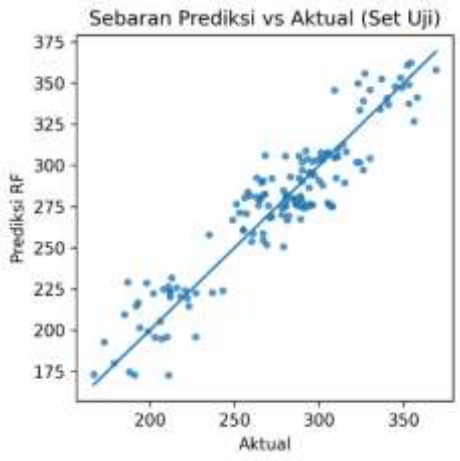
Kesesuaian profil prediksi *Random Forest* terhadap observasi aktual pada periode uji ditinjau melalui kurva deret waktu. Gambar 2 menampilkan kurva aktual dan prediksi *Random Forest* pada *holdout* uji. Pola puncak dan lembah umumnya terekam dengan baik, serta deviasi prediksi relatif terkendali pada hari-hari dengan perubahan moderat. Hal ini konsisten dengan capaian $R^2 \approx 0.873$ pada Tabel 2, yang menunjukkan model mampu menjelaskan proporsi variasi target yang tinggi pada periode uji.



Gambar 2. Aktual vs Prediksi *Random Forest* (Periode Uji)

Sebaran Prediksi terhadap Aktual

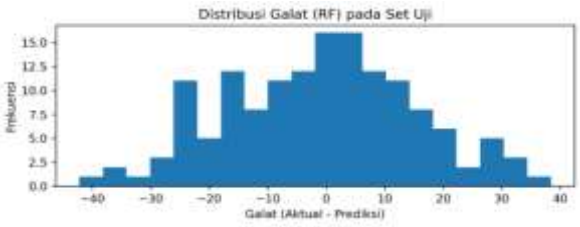
Hubungan antara prediksi dan aktual pada *holdout* uji divisualisasikan pada Gambar 3. Sebaran titik yang mendekati garis identitas ($y = x$) memperlihatkan tidak adanya bias sistematis yang menonjol pada rentang nilai utama. Titik yang berjarak lebih besar dari garis identitas mengindikasikan hari-hari dengan perubahan mendadak, yang secara alami lebih menantang bagi model deret waktu berbasis fitur kalender dan sinyal historis jangka pendek.



Gambar 3. *Scatter* Prediksi dan Aktual Set Uji dengan Garis Identitas

Distribusi Galat dan Indikasi Bias

Untuk menilai potensi bias sistematis, Gambar 4 menampilkan histogram *residuals* pada *holdout* uji. Sebaran galat berpusat di sekitar nol dan tidak menunjukkan ekor ekstrem yang panjang. Hal ini mengindikasikan tidak adanya bias sistematis besar pada periode uji dan konsisten dengan kinerja MAPE dan R^2 pada Tabel 3. Pemeriksaan terhadap beberapa hari dengan galat relatif besar menunjukkan kaitan dengan perubahan mendadak yang tidak dapat direpresentasikan oleh indikator kalender harian.

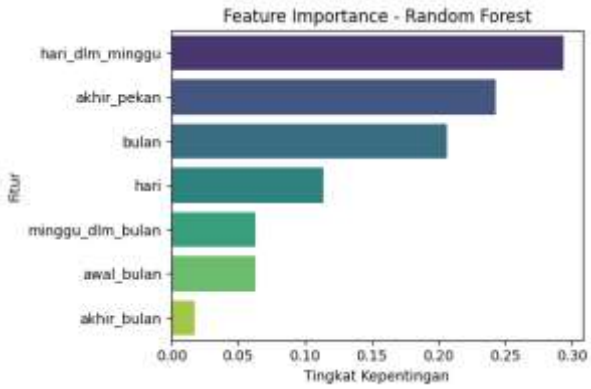


Gambar 4. Distribusi Galat (*Residuals*) Model *RF* pada Set Uji

Pentingnya Fitur (*Feature Importance*)

Kontribusi relatif fitur pada *Random Forest* ditunjukkan pada Gambar 5. Indikator hari dalam minggu (*hari_dlm_minggu*) serta akhir pekan (*akhir_pekan*) muncul dominan, diikuti bulan dan hari. Atribut minggu dalam bulan, awal bulan, dan akhir bulan turut berkontribusi meskipun lebih kecil. Pola ini selaras dengan dinamika operasional layanan publik yang dipengaruhi kalender kerja, perbedaan hari kerja dan akhir pekan, serta siklus awal dan akhir bulan.

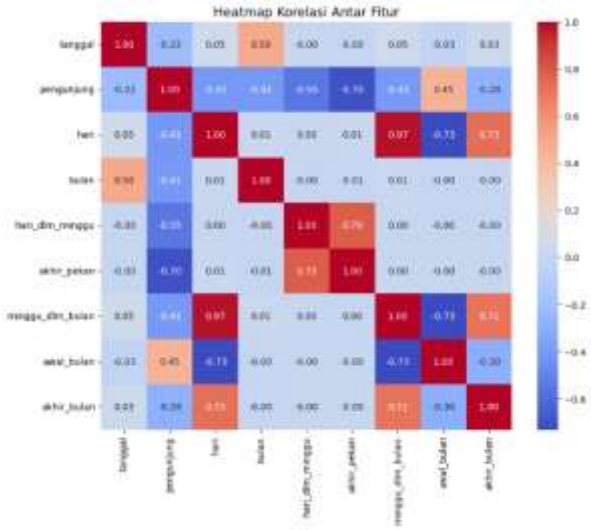
DOI : https://doi.org/10.52158/rjti.v4i3.4042



Gambar 5. Pentingnya fitur pada model *Random Forest*

Analisis Korelasi Antar Fitur

Gambar 6 menyajikan *heatmap* korelasi antar variabel fitur kalender dan target. Terlihat bahwa hari dan minggu_dlm_bulan memiliki korelasi sangat tinggi (≈ 0,97), demikian pula awal_bulan berkorelasi negatif kuat dengan minggu_dlm_bulan (≈ −0,73). Hal ini mengindikasikan adanya redundansi struktural di antara fitur-fitur kalender yang menghitung posisi hari dalam siklus bulan. Korelasi tinggi (≈ 0,79) juga muncul antara hari_dlm_minggu dan indikator akhir_pekan, yang wajar karena keduanya merepresentasikan dimensi yang sama (periodisitas mingguan).



Gambar 6. *Heatmap* Korelasi Antar Fitur

Terhadap variabel target, tampak korelasi negatif yang bermakna pada akhir_pekan (sekitar −0,70) dan hari_dlm_minggu (sekitar −0,55), yang menandakan rata-rata kunjungan lebih rendah di akhir pekan dalam *sample* ini. Sebaliknya, awal_bulan menunjukkan korelasi positif sedang (≈ 0,45), sementara bulan dan hari bernilai negatif sedang (sekitar −0,41). Temuan ini konsisten dengan *feature importance* pada *Random Forest* yang menempatkan indikator mingguan atau akhir pekan sebagai penentu utama (Gambar 5). Perlu dicatat bahwa kolom tanggal tidak digunakan sebagai prediktor dalam pemodelan, korelasinya terhadap fitur kalender terutama mencerminkan representasi waktu, bukan sinyal kausal langsung. Karena *Random Forest* relatif tahan terhadap multikolinearitas, redundansi fitur tidak merusak akurasi, tetapi dapat "membagi" bobot pentingnya fitur pada pasangan yang saling berkorelasi. Oleh karena itu, interpretasi *importance* dilakukan dengan kehati-hatian, dan *heatmap* ini membantu mengonfirmasi bahwa pola mingguan/kalender memang dominan pada data yang diteliti.

Implikasi Operasional

MAPE sekitar 5,00% pada *holdout* uji dan kesalahan absolut rata-rata (MAE) sekitar 13 pengunjung per hari menunjukkan tingkat presisi yang memadai untuk mendukung keputusan harian, seperti penjadwalan petugas dan pengaturan jam layanan. Mengingat fitur yang digunakan hanya berasal dari kalender, penerapan di lapangan relatif sederhana tanpa integrasi data eksternal yang kompleks. Perbandingan dengan dua *baseline* operasional (*naïve lag-1* dan rata-rata *day-of-week*) juga menunjukkan bahwa model ini mengurangi kesalahan prediksi lebih dari separuh, menegaskan bahwa pengambilan keputusan berbasis *machine learning* memiliki ketepatan yang lebih baik meskipun tanpa kompleksitas data yang tinggi.

Meskipun model *Random Forest* yang diusulkan menunjukkan kinerja yang kuat ($R^2 \approx 0.873$), penting untuk memahami keterbatasannya. Seperti yang telah diindikasikan pada analisis galat (Gambar 3 dan Gambar 4), keterbatasan utama model ini adalah sensitivitasnya terhadap peristiwa tak terduga (anomali) yang tidak terekam oleh fitur kalender. Model ini tidak mampu memprediksi lonjakan atau penurunan drastis yang disebabkan oleh faktor-faktor non-kalender, seperti: (1) Gangguan teknis, (2) Perubahan kebijakan operasional mendadak, atau (3) Hari libur khusus yang tidak teratur atau peristiwa sosial berskala besar yang tidak tercatat dalam data historis. Oleh karena itu, hari-hari dengan lonjakan tak biasa yang memunculkan galat lebih besar pada Gambar 3 dan Gambar 4 mengindikasikan ruang peningkatan akurasi apabila suatu saat integrasi fitur eksogen diperkenankan.

Sebagai pembanding konseptual, algoritma lain seperti *Gradient Boosting* dan *XGBoost* memang dikenal sangat kompetitif dan berpotensi memberikan peningkatan akurasi [24]. Namun, pemilihan *Random Forest* dalam penelitian ini didasarkan pada relevansi praktisnya untuk implementasi di lingkungan layanan publik dengan sumber daya komputasi terbatas. Fokus utama penelitian adalah membuktikan kelayakan bahwa model yang relatif sederhana, efisien, dan mudah direplikasi (seperti *Random Forest*) sudah mampu mengalahkan *baseline* operasional secara signifikan (reduksi *error* lebih dari 50%). Selain itu, *Random Forest* menawarkan

DOI : https://doi.org/10.52158/rjti.v4i3.4042

keunggulan interpretasi *feature importance* (Gambar 5) yang penting bagi pemangku kepentingan non-teknis, yang sejalan dengan tujuan penelitian. Dengan demikian, hasil yang dicapai menunjukkan bahwa bahkan tanpa kompleksitas model *boosting*, performa prediksi yang dihasilkan telah cukup kuat dan operasional untuk mendukung pengambilan keputusan berbasis data di lapangan.

4. Kesimpulan

Penelitian ini menunjukkan bahwa model *Random Forest* berbasis fitur kalender sederhana mampu menghasilkan prediksi volume pengunjung harian yang kompetitif pada sektor layanan publik. Dengan pemisahan berbasis waktu sekitar 80 persen pelatihan dan 20 persen *holdout* uji, model mencapai $R^2 = 0.873$, RMSE = 16.328, MAE = 13.103, serta MAPE sekitar 5,00%, dan terbukti melampaui dua *baseline* operasional, yaitu *naïve lag*-1 dan rata-rata *day-of-week*. Analisis *feature importance* mengindikasikan bahwa indikator hari dalam minggu dan akhir pekan merupakan determinan utama, selaras dengan pola musiman mingguan pada layanan publik. Implikasi kebijakan yang lebih konkret dari temuan ini adalah potensinya untuk menjadi fondasi bagi sistem perencanaan dinamis berbasis data (*data-driven planning system*) di instansi pemerintahan. Keunggulan pendekatan ini terletak pada biaya implementasi yang rendah, kemudahan replikasi, dan kesiapan operasional karena seluruh fitur bersumber dari informasi tanggal yang tersedia langsung. Keterbatasannya mencakup sensitivitas terhadap perubahan mendadak yang tidak tertangkap oleh fitur kalender harian, ruang lingkup fitur yang masih minimal, serta fokus horizon pada *one-step-ahead*.

Untuk pengembangan ke depan, disarankan pengayaan variabel eksogen (misalnya hari libur, promosi, cuaca, dan agenda lokal), perluasan skema fitur historis yang bebas *leakage* (lag musiman dan rata-rata bergerak), pembandingan dengan metode berbasis *gradient boosting*, serta penyediaan estimasi ketidakpastian (misalnya regresi kuantil atau interval prediksi). Selain itu, penyiapan alur penerapan yang mencakup penyimpanan model, *inference* harian, dasbor operasional, serta pemantauan kinerja berbasis MAE/MAPE bergulir untuk mendeteksi *drift*, perlu dikembangkan sebagai sistem prediktif jangka panjang. Terakhir, implementasi dan validasi eksternal pada unit layanan lain atau rentang tahun yang lebih panjang juga direkomendasikan agar generalisasi dan nilai kebijakan model semakin kuat.

Daftar Rujukan

[1] United Nations, "UN E-Government Survey 2024: Accelerating Digital Transformation for Sustainable Development, with the addendum on Artificial Intelligence," New York, 2024.

[2] Organisation for Economic Co-operation and Development (OECD), "2023 OECD Digital Government Index: Results and key findings," OECD Public Governance Policy Papers, No. 44, OECD Publishing, Paris, 2024.

[3] Organisation for Economic Co-operation and Development (OECD), "Governing with Artificial Intelligence: The State of Play and Way Forward in Core Government Functions," OECD Publishing, Paris, 2025.

[4] R. Madan and M. Ashok, "AI adoption and diffusion in public administration: A systematic literature review and future research agenda," Government Information Quarterly, vol. 40, no.1, 2023.

[5] F. Selten and B. Klievink, "Organizing public sector AI adoption: Navigating between separation and integration," Government Information Quarterly, vol. 41, no. 1, 2024.

[6] T. Haesevoets, B. Verschuere, and A. Roets, "AI adoption in public administration: Perspectives of public sector managers and public sector non-managerial employees," Government Information Quarterly, vol. 42, no. 2, 2025.

[7] E. Fatmawati, "Artificial Intelligence In Public Administration: Governance, Ethics, And Decision-Making," JV, vol. 17, no. 2, pp. 37-48, Aug. 2025.

[8] M. Mitzenmacher and R. Shahout, "Queueing, Predictions, and Large Language Models: Challenges and Open Problems," Stochastic Systems, vol. 15, no. 3, pp. 195-219, 2025, doi: 10.1287/stsy.2025.0106.

[9] F. R. di Torrepadula, E. V Napolitano, S. Di Martino, and N. Mazzocca, "Machine Learning for public transportation demand prediction: A Systematic Literature Review," Engineering Applications of Artificial Intelligence, vol. 137, 2024, doi: 10.1016/j.engappai.2024.109166.

[10] R. P. Munggaran, M. Nurmalasari, H. Hosizah, and D. Krismawati, "Prediksi Waktu Tunggu Pelayanan Pasien Rawat Jalan dengan Algoritma Random Forest: Predicting Outpatient Service Waiting Times with Random Forest Algorithm", MALCOM, vol. 5, no. 1, pp. 35-40, Nov. 2024.

[11] S. N. Windrasari, H. Margono, and Y. A. N. S. Putra, "Predictive Sales Analysis in Coffee Shops Using the Random Forest Algorithm,", MALCOM, vol. 5, no. 3, pp. 1000-1011, Jul. 2025.

[12] M. S. Zeb, M. A. Khan, M. M. H. Khattak, S. Ud-Din, M. F. Habib, and M. Z. Khan, "Forecasting public transit ridership amidst COVID-19: a machine learning approach," Public Transport, vol. 17, pp. 391–420, 2024, doi: 10.1007/s12469-024-00368-5.

[13] Å. Jevinger, C. Zhao, J. A. Persson, and P. Davidsson, "Artificial intelligence for improving public transport: a mapping study," Public Transport, vol. 16, pp. 99–158, 2024, doi: 10.1007/s12469-023-00334-7.

[14] H. Almukhalfi, A. Noor, and T. H. Noor "Traffic management approaches using machine learning and deep learning techniques: A survey," Engineering Applications of Artificial Intelligence, vol. 133, 2024, doi: 10.1016/j.engappai.2024.108147.

[15] B. M. Porto and F. S. Fogliatto, "Enhanced forecasting of emergency department patient arrivals using feature engineering approach and machine learning," BMC Med Inform Decis Mak, vol. 24, no. 377, 2024, doi: 10.1186/s12911-024-02788-6.

[16] C. Brossard, C. Goetz, P. Catoire, L. Cipolat, C. Guyeux, C. G. Jardine, M. Akplogan, and L. A. Vuillaume, "Predicting emergency department admissions using a machine-learning algorithm: a proof of concept with retrospective study," BMC Emerg Med, vol. 25, no. 3, 2025, doi: 10.1186/s12873-024-01141-4.

[17] L. Zhou, Q. Zhu, Q. Chen, P. Wang, and H. Huang, "Predicting hospital outpatient volume using XGBoost: a machine learning approach," Scientific reports, vol. 15, no. 1, 2025, doi: 10.1038/s41598-025-01265-y.

[18] C. Deina, F. S. Fogliatto, G. J. C. da Silveira, and M. J. Anzanello, "Decision analysis framework for predicting no-shows to appointments using machine learning algorithms,"

BMC Health Serv Res, vol. 24, no. 37, 2024, doi: 10.1186/s12913-023-10418-6.

[19] T. Oikonomidi, G. Norman, L. McGarrigle, J. Stokes, S. N. van der Veer, and D. Dowding, "Predictive model-based interventions to reduce outpatient no-shows: a rapid systematic review," Journal of the American Medical Informatics Association : JAMIA, vol. 30, no.3, pp. 559–569, 2023, doi: 10.1093/jamia/ocac242.

[20] S. Ryu, S. Jung, G. Kim, and S. Lee, "Visitor Number Prediction for Daegwallyeong Forest Trail Using Machine Learning," Sustainability, vol. 17, no. 13, 2025, doi: 10.3390/su17136061.

[21] M. Lu and Q. Xie, "A Novel Approach for Spatially Controllable High-Frequency Forecasts of Park Visitation Integrating Attention-Based Deep Learning Methods and Location-Based Services," ISPRS International Journal of Geo-Information, vol. 12, no.3, 2023, doi: 10.3390/ijgi12030098.

[22] Lukoseviciute G, Galanakis P, Martín-Nieto C, Wells G. K. Madden, G. Lukoseviciute, E. Ramsey, T. Panagopoulos, and J. Condell, "Forecasting daily foot traffic in recreational trails using machine learning." Journal of Outdoor Recreation and Tourism, 2023 doi: 10.1016/j.jort.2023.100701 ;43:100701.

[23] S. Uddin and H. Lu, "Confirming the statistically significant superiority of tree-based machine learning algorithms over their counterparts for tabular data," PLoS One, 2024, doi: 10.1371/journal.pone.0301541.

[24] L. Grinsztajn, E. Oyallon, and G. Varoquaux, "Why do tree-based models still outperform deep learning on typical tabular data?," in Proc. 36th Int. Conf. Neural Inf. Process. Syst. (NIPS '22), 2022, Art. no. 37, pp. 507–520.

[25] Z. Sadeghi, R. Alizadehsani et al., "A review of Explainable Artificial Intelligence in healthcare," Computers and Electrical Engineering, vol. 118, Aug. 2024, doi: 10.1016/j.compeleceng.2024.109370.

[26] H. Shiraishi, T. Nakamura, and R. Shibuki, "Time series quantile regression using random forests," J Time Ser Anal, vol. 45, no. 4, pp. 639–659, 2024, doi: 10.1111/jtsa.12731.

[27] R. J. Hyndman and G. Athanasopoulos, "Forecasting: Principles and Practice (3rd ed.)," 3rd ed. Melbourne: OTexts, 2021.

[28] A. D. Brabandere, T. O. D. Beéck, K. Hendrickx, W. Meert, and J. Davis, "TSFuse: automated feature construction for multiple time series data," Mach Learn vol. 113, pp. 5001–5056, 2024, doi: 10.1007/s10994-021-06096-2.

[29] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, "M5 accuracy competition: Results, findings, and conclusions," International Journal of Forecasting, vol. 38, no. 4, pp. 1346-1364, 2022, doi: 10.1016/j.ijforecast.2021.11.013.

[30] A. Alroomi, G. Karamatzanis, K. Nikolopoulos, A. Tilba, and S. Xiao, "Fathoming empirical forecasting competitions' winners," International Journal of Forecasting, vol, 38, no. 4, pp. 1519-1525, 2022, doi: 10.1016/j.ijforecast.2022.03.010.

[31] F. Zito, V. Cutello, and M. Pavone, "Data-driven forecasting and its role in enhanced decision-making," Engineering Applications of Artificial Intelligence, vol. 154, 2025, doi: 10.1016/j.engappai.2025.110934.

[32] Tommaso Proietti, Diego J. Pedregal, "Seasonality in High Frequency Time Series," Econometrics and Statistics, vol. 27, pp. 62-82, 2023, doi: 10.1016/j.ecosta.2022.02.001.

[33] S. B. Chudo, and G. Terdik, "Modeling and Forecasting Time-Series Data with Multiple Seasonal Periods Using Periodograms," Econometrics, vol. 13, no. 2, 2025, doi: 10.3390/econometrics13020014

[34] F. N. Farida, A. Faqih, and S. E. Permana, "Penerapan Model Prediksi Penjualan pada Usaha Rumah Makan Menggunakan Algoritma Random Forest," JATI (Jurnal Mahasiswa Teknik Informatika), vol. 9, no. 4, pp. 5895-5902, 2025, doi: 10.36040/jati.v9i4.13912.

[35] A. Sari, M. Arifin, and E. Darmanto, "Prediksi Kebutuhan Stok Barang Menggunakan Algoritma Random Forest Untuk Meningkatkan Efisiensi Penjualan," JOISIE (Journal Of Information Systems And Informatics Engineering), vol. 9, no. 2, pp. 339-351, 2025, doi: 10.35145/joisie.v9i2.5154.

[36] M. S. Efendi, S. Sarwido, and A. K. Zyen, "Penerapan Algoritma Random Forest Untuk Prediksi Penjualan Dan Sistem Persediaan Produk, " Resolusi : Rekayasa Teknik Informatika dan Informasi, vol. 5, no. 1, pp. 12-20 , 2024, doi: 10.30865/resolusi.v5i1.2149.