

Predicting Daily Patient Visits at Community Health Centers Using Machine Learning Algorithms to Optimize Medical Staff Scheduling

Zahza Ivana Putri ¹⁾, Edy Hermansyah ²⁾, Dimas Aulia Trianggana ³⁾

¹⁾ Study Program of Informatics, Faculty of Computer Science, Universitas Dehasen Bengkulu

^{2,3)} Department of Informatics, Faculty of Computer Science, Universitas Dehasen Bengkulu

Email: ¹⁾ zazaivana7@gmail.com

How to Cite :

Putri. Z.I., Hermansyah. E & Trianggana.D.A (2026). Prediction Of Daily Patient Visits At A Community Health Center Using A Machine Learning Algorithm For Optimizing Medical Staff Scheduling. Jurnal Media Computer Science, 5(4).

ARTICLE HISTORY

Received [31 Agustus 2026]

Revised [25 Oktober 2026]

Accepted [28 Oktober 2026]

KEYWORDS

Patient Visits, Machine Learning, Random Forest .

This is an open access article under the
[CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license



ABSTRAK

Pelayanan kesehatan yang optimal di tingkat fasilitas kesehatan tingkat pertama sangat bergantung pada kesiapan sumber daya manusia. Di Puskesmas Lempuing Bengkulu, terjadi ketidakseimbangan operasional yang signifikan akibat fluktuasi jumlah kunjungan harian pasien yang tidak menentu. Penjadwalan tenaga medis, selama ini masih dilakukan secara konvensional tanpa mempertimbangkan tren kedatangan pasien di masa mendatang. Hal ini berdampak pada antrean yang panjang pada hari-hari tertentu, sekaligus menyebabkan inefisiensi tenaga kerja pada saat tingkat kunjungan sedang rendah.. Penelitian ini menggunakan metodologi Cross-Industry Standard Process for Data Mining (CRISP-DM). Pendekatan pemodelan dilakukan dengan menerapkan algoritma Random Forest. Proses komputasi memanfaatkan data historis kunjungan harian pasien selama periode tiga bulan terakhir, yakni dari bulan Oktober hingga Desember 2025. Selain membangun model prediksi secara komputasional, penelitian ini juga mengintegrasikan hasil forecasting dengan formula Standar Beban Kerja (SBK). Berdasarkan hasil pengujian model Random Forest menunjukkan performa evaluasi dengan nilai Mean Absolute Error (MAE) sebesar 7,27 pasien per hari dan Root Mean Squared Error (RMSE) sebesar 10,25 pasien. Nilai koefisien determinasi yang didapatkan adalah 37,10%, yang secara wajar mengindikasikan tingkat kompleksitas dan sifat data historis kesehatan yang sangat stokastik serta rentan dipengaruhi oleh faktor eksternal acak. Analisis kepentingan fitur mengungkapkan bahwa variabel Kunjungan H-7 dan H-1 merupakan parameter yang paling dominan dalam memengaruhi lonjakan jumlah pasien. sistem ini secara efektif mampu mentransformasikan angka prediksi digital menjadi rekomendasi alokasi tenaga medis, yakni berkisar antara 1 hingga 3 setiap harinya.

ABSTRACT

Optimal health care at primary-level health facilities depends heavily on the readiness of human resources. At the Lempuing Community Health Center in Bengkulu, there is a significant operational imbalance due to unpredictable fluctuations in the number of daily patient visits. Medical staff scheduling has thus far been conducted conventionally without considering future trends in patient arrivals. This results in long lines on certain days, while also causing labor inefficiency during periods of low patient volume. This study employs the Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology. The modeling approach utilizes the Random Forest algorithm. The computational process leverages historical data on daily patient visits over the past three months, specifically from October through December 2025. In addition to building a computational prediction model, this study also integrates the forecasting results with the Standard Workload (SBK)

formula. Based on the results of the Random Forest model testing, the evaluation performance showed a Mean Absolute Error (MAE) of 7.27 patients per day and a Root Mean Squared Error (RMSE) of 10.25 patients. The coefficient of determination obtained was 37.10%, which reasonably indicates the level of complexity and the highly stochastic nature of historical health data, which is susceptible to random external factors. Feature importance analysis revealed that the H-7 and H-1 Visit variables are the most dominant parameters influencing spikes in patient numbers. The system is effectively capable of transforming digital prediction figures into recommendations for medical staff allocation, ranging from 1 to 3 per day.

PENDAHULUAN

Perkembangan teknologi informasi dan analisis data mendorong fasilitas kesehatan untuk memanfaatkan rekam data operasional sebagai dasar pengambilan keputusan. Pada layanan kesehatan tingkat pertama, jumlah pasien tidak selalu stabil dari satu hari ke hari berikutnya. Fluktuasi tersebut berpengaruh langsung terhadap beban pelayanan, panjang antrian, dan pemanfaatan tenaga medis. Puskesmas Lempuing Kota Bengkulu telah memiliki pencatatan kunjungan harian, tetapi data tersebut terutama dimanfaatkan sebagai arsip dan pelaporan. Penjadwalan tenaga medis masih banyak bergantung pada pengalaman serta perkiraan, sehingga informasi historis belum sepenuhnya digunakan untuk mengantisipasi pola kedatangan pasien.

Kondisi operasional tersebut menimbulkan persoalan ketika jumlah petugas yang dijadwalkan tidak sebanding dengan volume pasien. Observasi awal pada skripsi menunjukkan adanya hari dengan beban pendaftaran yang tinggi sementara pelayanan ditangani oleh jumlah petugas yang terbatas. Ketidaksiuaian ini berpotensi memperpanjang waktu tunggu pada hari padat, sedangkan penempatan personel yang sama pada hari dengan kunjungan lebih rendah dapat menimbulkan inefisiensi. Karena itu, kebutuhan utama bukan sekadar mengetahui jumlah kunjungan yang telah terjadi, tetapi memperkirakan jumlah kunjungan berikutnya sehingga alokasi personel dapat disiapkan sebelum layanan dimulai.

Machine Learning menawarkan pendekatan prediktif yang dapat mempelajari pola dari data historis. Asaury et al. (2025) menunjukkan pemanfaatan Random Forest untuk memprediksi jumlah pasien masuk rumah sakit, sedangkan Nugroho et al. (2024) menggunakan regresi linier sederhana untuk memperkirakan kunjungan pasien rawat jalan. Pranajaya dan Susanto (2025) juga memperlihatkan bahwa Random Forest dapat dioptimalkan pada konteks kesehatan melalui pengolahan data dan pencarian parameter. Temuan-temuan tersebut menguatkan potensi pemodelan prediktif, tetapi objek, karakter data, dan orientasi implementasinya berbeda dari kebutuhan operasional layanan primer.

Gap penelitian terlihat dari studi terdahulu yang umumnya memisahkan prediksi kunjungan dan penjadwalan tenaga. Cubillas et al. (2022) membangun prediksi arus pasien pada pusat layanan kesehatan primer, Selya et al. (2021) berfokus pada kunjungan medis tidak terencana pada tingkat individu, Erkamp et al. (2021) dan Fan et al. (2022) mempelajari volume kedatangan pada unit gawat darurat, sedangkan Koruca et al. (2023) berorientasi pada optimasi jadwal staf. Belum tampak integrasi langsung antara prediksi volume kunjungan harian berbasis Random Forest dan konversi hasil prediksi menjadi jumlah tenaga yang diperlukan dengan ukuran beban kerja pada konteks Puskesmas Lempuing. Kebaruan penelitian ini terletak pada integrasi dua lapis keputusan. Lapisan pertama menghasilkan estimasi jumlah pasien harian dari fitur temporal dan histori kunjungan menggunakan Random Forest Regressor. Lapisan kedua menerjemahkan estimasi tersebut ke dalam kebutuhan tenaga melalui Standar Beban Kerja, sehingga keluaran model tidak berhenti pada angka prediksi, tetapi dapat digunakan sebagai rekomendasi operasional. Dengan ruang lingkup data Oktober–Desember 2025, penelitian difokuskan pada prediksi dan evaluasi model, bukan implementasi sistem informasi penuh.

LANDASAN TEORI

Penelitian Terdahulu

Penelitian terdahulu memperlihatkan bahwa pemodelan kunjungan pasien dan optimasi tenaga kesehatan telah dikaji melalui pendekatan yang beragam. Namun, fokus prediksi, unit layanan, variabel, dan bentuk pemanfaatan hasil model belum sepenuhnya sama dengan penelitian ini. Ringkasan studi yang menjadi pijakan gap penelitian disajikan pada Tabel 1.

Tabel 1. Ringkasan Penelitian Terdahulu

No	Penulis	Metode/Objek	Temuan dan Kesenjangan
1	Cubillas et al. (2022)	General Linear Model dan Support Vector Machine; primary healthcare centres	Kesalahan absolut prediksi dilaporkan rendah dan hasil dikembangkan menuju sistem pakar, tetapi tidak menggunakan Random Forest dan tidak mengintegrasikan prediksi dengan kebutuhan tenaga.
2	Selya et al. (2021)	Support Vector Machine dan Logistic Regression; pasien diabetes	Berfokus pada prediksi kunjungan medis tidak terencana pada level individu, bukan volume kunjungan fasilitas kesehatan.
3	Erkamp et al. (2021)	Machine Learning dan time-series forecasting; emergency department	Model berbasis Machine Learning meningkatkan kemampuan prediksi, tetapi konteksnya rumah sakit pendidikan dan bukan layanan primer.
4	Koruca et al. (2023)	Mathematical optimization dan staff scheduling	Menghasilkan jadwal staf yang lebih sesuai beban kerja, namun kebutuhan tenaga tidak diturunkan dari prediksi kunjungan berbasis Machine Learning.
5	Fan et al. (2022)	Ensemble learning dan forecasting; emergency department	Model ensemble stabil dalam menghadapi fluktuasi kedatangan, tetapi hasil prediksi tidak diarahkan pada penjadwalan tenaga medis.

Random Forest dan Prediksi Kunjungan

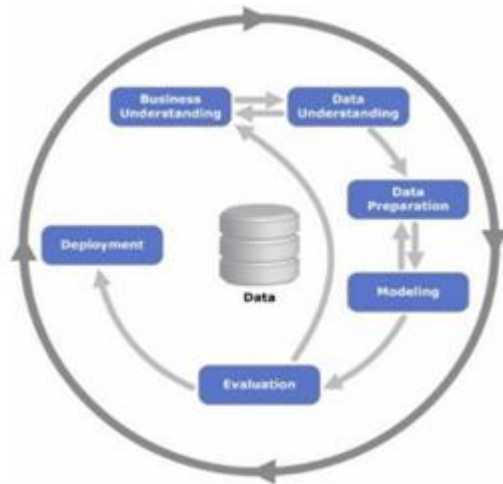
Random Forest merupakan metode ensemble yang menggabungkan banyak pohon keputusan untuk membentuk hasil prediksi yang lebih stabil. Pada masalah regresi, keluaran akhir diperoleh dari agregasi prediksi pohon-pohon yang dibangun dari sampel dan subset fitur yang dipilih secara acak. Karakter tersebut membantu menekan ketergantungan pada satu pohon dan memungkinkan model menangkap hubungan nonlinier pada data. Sultan et al. (2025) menjelaskan Random Forest sebagai pendekatan ensemble yang digunakan untuk meningkatkan kemampuan prediksi dengan memanfaatkan koleksi pohon keputusan.

Dalam konteks kunjungan pasien, data memiliki karakter runtun waktu sehingga informasi tanggal dan riwayat kedatangan menjadi penting. Penelitian ini membentuk fitur hari dalam minggu, kunjungan satu hari sebelumnya (H-1), kunjungan pada hari yang sama satu minggu sebelumnya (H-7), dan penanda akhir pekan. Fitur tersebut digunakan untuk membantu model mengenali kecenderungan jangka pendek dan pola berulang mingguan tanpa mengubah target utama berupa total kunjungan harian.

Cross-Industry Standard Process for Data Mining dan Standar Beban Kerja

Alur penelitian disusun dengan kerangka Cross-Industry Standard Process for Data Mining yang mencakup business understanding, data understanding, data preparation, modeling, evaluation, dan deployment. Kerangka ini dipilih karena memberi urutan kerja yang sistematis dari pemahaman masalah hingga pemanfaatan hasil model (Falah et al., 2025). Pada penelitian ini,

deployment dimaknai sebagai tahap konversi hasil prediksi menjadi rekomendasi operasional, bukan penerapan aplikasi produksi secara penuh.



Gambar 1. Tahapan Cross-Industry Standard Process for Data Mining

Agar prediksi memiliki makna praktis, hasil model dihubungkan dengan analisis beban kerja. Perhitungan kebutuhan tenaga dilakukan dengan membandingkan estimasi jumlah pasien dengan kapasitas pelayanan per petugas. Pendekatan ini sejalan dengan penggunaan ukuran beban kerja untuk memperkirakan kebutuhan personel pada layanan kesehatan (Tuzzakiyah et al., 2022). Dalam data penelitian, waktu pelayanan efektif adalah 180 menit per hari dan durasi pelayanan sekitar lima menit per pasien, sehingga kapasitas ideal yang digunakan pada tahap integrasi adalah 36 pasien per petugas per hari.

METODE PENELITIAN

Penelitian menggunakan pendekatan kuantitatif dengan subjek analisis berupa data historis kunjungan harian pasien di Puskesmas Lempuing Kota Bengkulu. Kegiatan penelitian berlangsung pada November 2025 sampai Mei 2026. Sumber data utama berupa dokumen kunjungan harian yang dikumpulkan untuk periode Oktober, November, dan Desember 2025. Observasi digunakan untuk memahami kondisi pencatatan dan penjadwalan aktual, studi dokumentasi digunakan untuk memperoleh data kunjungan, dan studi pustaka digunakan untuk mendukung pemilihan metode serta interpretasi hasil.

Pengolahan mengikuti enam tahap Cross-Industry Standard Process for Data Mining. Pada business understanding, masalah didefinisikan sebagai ketidakseimbangan antara volume pasien dan jumlah tenaga yang bertugas. Data understanding dilakukan dengan menelaah struktur berkas, atribut, kelengkapan data, missing value, dan kemungkinan outlier. Data preparation mencakup penggabungan dokumen bulanan, pembersihan kolom kosong atau tidak konsisten, konversi tanggal, serta pembentukan fitur temporal. Tahap modeling membangun Random Forest Regressor pada Google Colab menggunakan Python 3.11.

Pembagian data mempertahankan urutan waktu dan tidak menggunakan pengacakan acak. Data Oktober dan November 2025 digunakan sebagai data latih untuk mempelajari pola kunjungan, sedangkan Desember 2025 digunakan sebagai data uji untuk mensimulasikan kondisi prediksi pada periode berikutnya. Strategi ini dipilih agar informasi masa depan tidak masuk ke proses pelatihan dan evaluasi tetap mencerminkan penggunaan model secara prospektif.

Fitur hasil rekayasa mencakup hari dalam minggu, kunjungan H-1, kunjungan H-7, dan status akhir pekan. Target model adalah jumlah total kunjungan pasien per hari. Kinerja model dievaluasi menggunakan Mean Absolute Error, Root Mean Squared Error, dan koefisien determinasi. Mean Absolute Error menggambarkan rata-rata selisih absolut antara nilai aktual dan prediksi, sedangkan Root Mean Squared Error memberi penalti lebih besar pada deviasi yang lebih tinggi (Ihzaniah et al., 1856 Zahza Ivana Putri, Edy Hermansyah, Dimas Aulia Trianggana : Predicting Daily Patient Visits at Community Health ...

2023). Koefisien determinasi digunakan untuk menunjukkan proporsi variasi target yang dapat dijelaskan oleh model pada data uji.

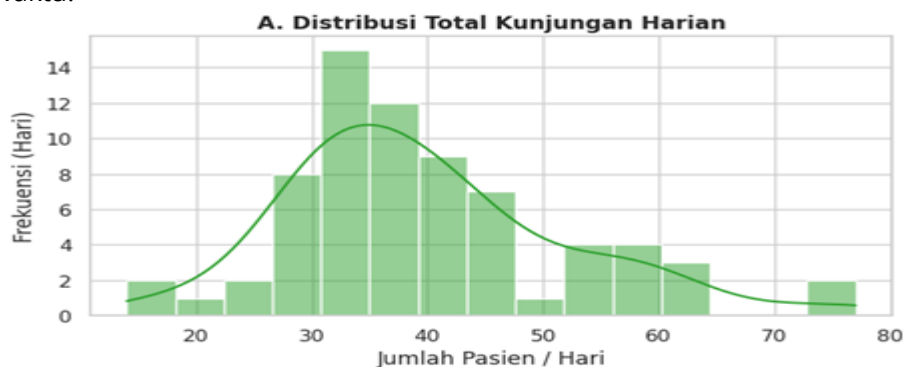
Tahap akhir mengintegrasikan prediksi dengan Standar Beban Kerja. Dengan kapasitas 36 pasien per petugas per hari, kebutuhan petugas dihitung dari jumlah kunjungan hasil prediksi dan dibulatkan ke atas agar kapasitas pelayanan tidak berada di bawah estimasi kebutuhan. Mekanisme tersebut menghasilkan rekomendasi personel dinamis pada rentang 1–3 petugas per hari. Penelitian dibatasi pada tahap analisis, prediksi, evaluasi, dan formulasi rekomendasi; sistem belum diimplementasikan sebagai aplikasi operasional penuh.

HASIL DAN PEMBAHASAN

Hasil

Proses implementasi dimulai dari penggabungan data kunjungan pasien BPJS dan umum selama tiga bulan. Dokumen sumber memiliki ketidakseragaman posisi kop, penomoran, dan kolom kosong. Tahap pembersihan pada Google Colab digunakan untuk mendeteksi baris utama, membuang kolom tanpa nama yang tidak digunakan, menyeragamkan struktur data, dan menghasilkan satu dataset yang dapat diproses secara konsisten. Setelah data dibersihkan, tanggal dikonversi ke format waktu dan fitur temporal dibentuk untuk kebutuhan model.

Exploratory Data Analysis menunjukkan adanya fluktuasi kunjungan harian, sehingga penggunaan satu jumlah petugas yang tetap untuk seluruh hari tidak sepenuhnya mencerminkan perubahan beban pelayanan. Distribusi data memperlihatkan bahwa sebagian besar hari memiliki volume kunjungan menengah, tetapi terdapat pula hari dengan volume lebih rendah dan lonjakan lebih tinggi. Variasi tersebut menjadi dasar perlunya model yang dapat memperhitungkan pola historis dan waktu.



Gambar 2. Distribusi Total Kunjungan Harian Pasien

Model Random Forest kemudian dilatih menggunakan data Oktober–November dan diuji pada data Desember. Secara visual, hasil prediksi mengikuti sebagian pola kenaikan dan penurunan kunjungan aktual, walaupun terdapat deviasi pada beberapa hari dengan perubahan tajam. Kondisi ini konsisten dengan karakter data pelayanan kesehatan yang dipengaruhi faktor eksternal yang tidak seluruhnya tersedia pada dataset.



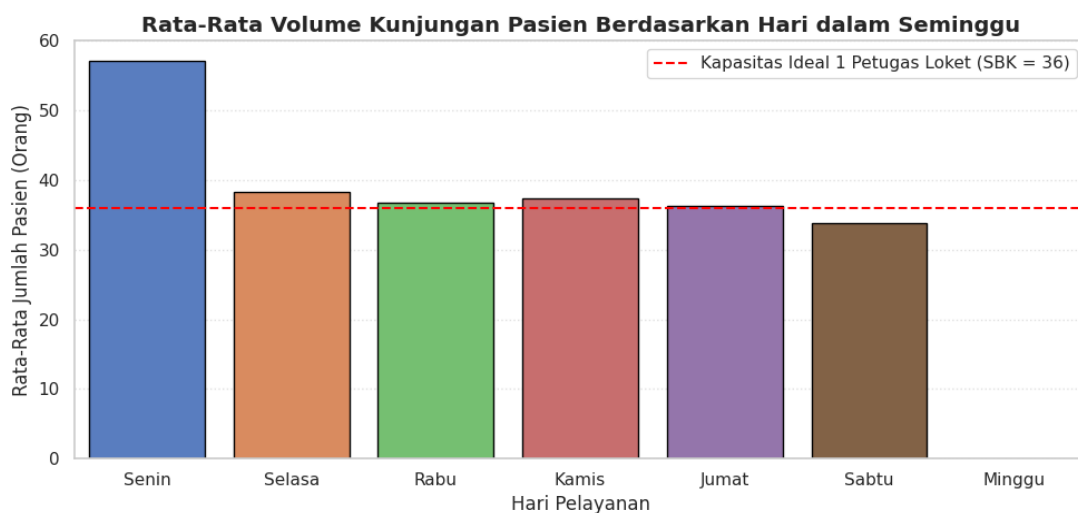
Gambar 3. Perbandingan Kunjungan Aktual dan Hasil Prediksi Random Forest pada Data Uji

Evaluasi kuantitatif menghasilkan Mean Absolute Error sebesar 7,27 pasien per hari, Root Mean Squared Error sebesar 10,25 pasien per hari, dan koefisien determinasi sebesar 0,3710 atau 37,10%. Nilai kesalahan menunjukkan bahwa selisih prediksi harian berada pada skala yang lebih kecil dibanding kapasitas pelayanan satu petugas yang digunakan dalam perhitungan Standar Beban Kerja, yaitu 36 pasien per hari. Ringkasan hasil evaluasi disajikan pada Tabel 2.

Tabel 2. Hasil Evaluasi Model Random Forest

No	Metrik	Hasil	Interpretasi dalam Penelitian
1	Mean Absolute Error	7,27 pasien/hari	Rata-rata selisih absolut antara prediksi dan kunjungan aktual sekitar tujuh pasien per hari.
2	Root Mean Squared Error	10,25 pasien/hari	Menunjukkan besarnya deviasi prediksi dengan penalti lebih besar pada kesalahan yang lebih tinggi.
3	Koefisien Determinasi	0,3710 (37,10%)	Model menangkap sebagian variasi struktural kunjungan, sementara variasi lain dipengaruhi faktor yang tidak tersedia pada data.

Analisis fitur pada skripsi menunjukkan bahwa informasi kunjungan historis H-7 dan H-1 memberikan kontribusi penting terhadap pembentukan prediksi. H-7 merepresentasikan pola pada hari yang sama di minggu sebelumnya, sedangkan H-1 menangkap perubahan jangka pendek dari hari sebelumnya. Status akhir pekan berkaitan dengan penurunan volume kunjungan karena jam pelayanan dan pola kedatangan berbeda menjelang hari libur. Temuan ini memperlihatkan bahwa struktur temporal merupakan komponen penting ketika data eksternal seperti cuaca, hari libur nasional, dan tren epidemiologi belum dimasukkan ke model. Rata-rata kunjungan berdasarkan hari pelayanan memperlihatkan beban paling tinggi pada Senin dan cenderung lebih rendah menuju akhir minggu. Garis kapasitas 36 pasien pada Gambar 4 menunjukkan batas pelayanan ideal satu petugas yang digunakan dalam integrasi Standar Beban Kerja. Pola ini menjelaskan mengapa jumlah tenaga yang sama setiap hari dapat menghasilkan kondisi yang kurang adaptif terhadap beban aktual.



Gambar 4. Rata-Rata Volume Kunjungan Pasien Berdasarkan Hari dan Kapasitas Ideal Satu Petugas

Dengan membagi jumlah pasien hasil prediksi terhadap kapasitas 36 pasien per petugas dan menerapkan pembulatan ke atas, sistem menghasilkan kebutuhan tenaga pada rentang 1–3 petugas. Rentang tersebut dimaksudkan untuk menjaga agar alokasi tenaga mengikuti estimasi beban harian dan menyediakan ruang kapasitas terhadap deviasi prediksi. Hasil integrasi dirangkum pada Tabel 3.

Tabel 3. Integrasi Prediksi dengan Standar Beban Kerja

Komponen	Nilai/Aturan	Implikasi Operasional
Waktu pelayanan efektif	180 menit/hari	Menjadi dasar kapasitas pelayanan harian.
Durasi pelayanan	5 menit/pasien	Menghasilkan kapasitas ideal 36 pasien per petugas per hari.
Kesalahan rata-rata model	Mean Absolute Error 7,27 pasien/hari	Deviasi rata-rata berada di bawah kapasitas satu petugas.
Kebutuhan hasil integrasi	1–3 petugas/hari	Jumlah petugas disesuaikan dengan prediksi dan dibulatkan ke atas.

Pembahasan

Hasil penelitian menunjukkan bahwa Random Forest dapat membentuk pola prediksi yang cukup dekat dengan data aktual pada sebagian besar hari uji, tetapi tidak menjelaskan seluruh variasi kunjungan. Koefisien determinasi 37,10% perlu dibaca bersama sifat data layanan kesehatan yang dipengaruhi faktor di luar dataset. Dalam penelitian ini, variabel cuaca, kalender libur, wabah atau tren penyakit, dan preferensi waktu pasien belum digunakan. Karena itu, nilai koefisien determinasi tidak diperlakukan sebagai satu-satunya ukuran kelayakan operasional; Mean Absolute Error dan hubungannya dengan kapasitas pelayanan menjadi indikator yang penting untuk menilai konsekuensi praktis kesalahan prediksi.

Temuan ini sejalan dengan arah penelitian Asaury et al. (2025) yang menunjukkan Random Forest dapat dimanfaatkan untuk prediksi jumlah pasien, namun konteks Puskesmas Lempuing menambahkan kebutuhan untuk menghubungkan prediksi dengan keputusan personel. Studi Cubillas et al. (2022), Erkamp et al. (2021), dan Fan et al. (2022) juga menegaskan bahwa data historis dan model prediktif dapat membantu memperkirakan arus pasien. Perbedaan penelitian ini terletak pada pemanfaatan keluaran regresi sebagai masukan langsung bagi formula kebutuhan tenaga, sehingga hasil analitik dihubungkan dengan proses manajerial harian.

Pola kunjungan mingguan memperkuat alasan penggunaan fitur temporal. Kunjungan yang lebih tinggi pada awal minggu dan kecenderungan menurun pada akhir minggu menunjukkan bahwa beban pelayanan tidak homogen. H-1 menyediakan konteks paling dekat dengan kondisi terkini, sedangkan H-7 membawa informasi siklus mingguan. Walaupun tidak seluruh perubahan dapat diterangkan oleh fitur tersebut, kombinasi histori dekat dan siklus mingguan memberi dasar yang lebih informatif dibanding penjadwalan yang hanya mengandalkan rekap bulanan atau jumlah personel tetap.

Integrasi dengan Standar Beban Kerja merupakan aspek yang membedakan hasil prediksi dari sekadar keluaran komputasional. Dengan kapasitas 36 pasien per petugas per hari, Mean Absolute Error 7,27 pasien masih berada di bawah kapasitas satu petugas. Pembulatan kebutuhan tenaga ke atas berfungsi sebagai buffer operasional ketika prediksi meleset beberapa pasien. Pada hari padat, rekomendasi dapat meningkat hingga tiga petugas, sedangkan pada hari dengan beban lebih rendah jumlah petugas dapat dikurangi dan personel lain dialihkan ke pekerjaan administrasi atau pelaporan. Prinsip ini sejalan dengan orientasi optimasi jadwal tenaga yang dibahas Koruca et al. (2023), tetapi kebutuhan jumlah staf pada penelitian ini terlebih dahulu dipicu oleh prediksi volume kunjungan.

Secara ilmiah, hasil ini menunjukkan bahwa nilai prediksi tidak harus berdiri sendiri. Ketika metrik error diinterpretasikan terhadap kapasitas pelayanan, kegunaan model dapat dinilai dari dampaknya pada keputusan. Novelty logis penelitian ini adalah jembatan antara forecasting dan workload-based staffing pada layanan primer. Model menghasilkan estimasi pasien, sedangkan Standar Beban Kerja mengubah estimasi tersebut menjadi unit keputusan yang dapat digunakan manajemen, yaitu jumlah petugas per hari. Dengan demikian, keluaran Machine Learning menjadi lebih mudah diterjemahkan ke kebutuhan operasional.

Meskipun demikian, model belum dapat dianggap sebagai penentu jadwal final tanpa evaluasi lanjutan. Dataset hanya mencakup tiga bulan dan belum memasukkan faktor eksternal yang dapat memengaruhi kedatangan pasien. Selain itu, penelitian belum membandingkan Random Forest

secara empiris dengan Gradient Boosting, XGBoost, CatBoost, atau model runtun waktu lain pada dataset yang sama. Keterbatasan tersebut membuka ruang pengembangan agar prediksi lebih stabil dan generalisasi model dapat diuji pada periode yang lebih panjang.

Keterkaitan Temuan dengan Studi Terdahulu

Perbandingan dengan penelitian terdahulu memperjelas posisi temuan. Cubillas et al. (2022) sama-sama menempatkan arus pasien pada layanan kesehatan primer sebagai objek prediksi, tetapi pendekatan tersebut menggunakan model regresi dan Support Vector Machine serta mengarah pada sistem pakar. Penelitian di Puskesmas Lempuing menggunakan Random Forest dan memanfaatkan hasil prediksi sebagai masukan untuk kebutuhan tenaga. Dengan demikian, kesamaan objek layanan primer tidak menghasilkan keluaran manajerial yang sama; penelitian ini menambahkan tahap konversi prediksi menjadi jumlah petugas.

Selya et al. (2021) memperlihatkan penggunaan Machine Learning untuk memprediksi kunjungan medis tidak terencana pada pasien diabetes. Fokus tersebut berada pada risiko kunjungan individu, sedangkan penelitian ini menempatkan volume kunjungan fasilitas sebagai target. Perbedaan unit analisis ini penting karena kebutuhan penjadwalan tidak ditentukan oleh risiko seorang pasien, melainkan oleh akumulasi jumlah pasien yang harus dilayani pada satu hari. Erkamp et al. (2021) dan Fan et al. (2022) menunjukkan relevansi model Machine Learning untuk forecasting kedatangan di emergency department, tetapi karakter unit rumah sakit dan layanan primer memiliki pola operasional yang berbeda.

Koruca et al. (2023) mendekati persoalan dari sisi penjadwalan tenaga dan menunjukkan bahwa pengaturan staf dapat disesuaikan dengan beban kerja. Penelitian ini melengkapi arah tersebut dengan menentukan beban harian terlebih dahulu melalui prediksi Random Forest. Hubungan antara dua tahap ini membuat keputusan tenaga tidak hanya bersifat optimasi matematis terhadap jadwal yang telah diketahui, tetapi dimulai dari estimasi kebutuhan layanan yang belum terjadi. Integrasi ini menjadi dasar agar rekomendasi 1–3 petugas dapat berubah mengikuti prediksi jumlah pasien.

Implikasi Operasional dan Validasi Penggunaan Model

Dalam skenario operasional, hasil prediksi dapat dihitung sebelum hari pelayanan dan kemudian dibandingkan dengan kapasitas 36 pasien per petugas. Jika estimasi kunjungan meningkat, jumlah petugas dapat ditambah dengan pembulatan ke atas agar kapasitas tidak lebih kecil dari beban yang diperkirakan. Sebaliknya, saat proyeksi menurun, sebagian tenaga dapat dialihkan ke pekerjaan administrasi atau pelaporan. Pola rata-rata kunjungan yang lebih tinggi pada awal minggu dan lebih rendah menjelang akhir minggu mendukung kebutuhan penjadwalan yang tidak seragam sepanjang minggu.

Mean Absolute Error sebesar 7,27 pasien per hari menjadi penting ketika dibaca bersama Standar Beban Kerja. Nilai tersebut lebih kecil daripada kapasitas 36 pasien yang digunakan untuk satu petugas per hari, sehingga deviasi rata-rata model masih berada di dalam satu unit kapasitas personel. Fungsi pembulatan ke atas berperan sebagai ruang pengaman ketika prediksi tidak sama dengan jumlah aktual. Namun, buffer tersebut tidak menghilangkan kebutuhan pemantauan karena pada hari dengan perubahan ekstrem, Root Mean Squared Error mencapai 10,25 pasien per hari dan dapat memengaruhi kedekatan antara estimasi dan realisasi.

Pemanfaatan model juga membutuhkan siklus evaluasi. Data aktual setiap hari dapat ditambahkan ke basis historis untuk menilai apakah pola H-1, H-7, dan hari pelayanan tetap relevan. Analisis kebutuhan nonfungsional pada skripsi mengarahkan sistem agar dapat menerima penambahan data baru dan melakukan pelatihan ulang secara berkala. Dengan cara ini, model tidak diperlakukan sebagai konfigurasi tetap, melainkan sebagai alat analitik yang dapat diperbarui ketika karakter kunjungan berubah.

Validasi berbasis urutan waktu pada penelitian ini memberi gambaran penggunaan prospektif karena data Oktober dan November digunakan untuk pelatihan dan Desember digunakan sebagai

pengujian. Strategi tersebut menjaga agar informasi periode uji tidak digunakan saat model belajar. Meskipun demikian, satu bulan pengujian belum cukup untuk menunjukkan kestabilan pada musim atau kondisi layanan yang berbeda. Oleh sebab itu, perluasan data lebih dari tiga bulan merupakan langkah penting sebelum rekomendasi diterapkan sebagai dasar utama penjadwalan.

Keterbatasan variabel juga memengaruhi interpretasi koefisien determinasi 37,10%. Perubahan cuaca, hari libur nasional, tren epidemiologi penyakit, dan faktor perilaku pasien tidak tersedia pada dataset, sehingga model hanya memanfaatkan struktur waktu dan histori kunjungan yang telah dibentuk. Penambahan fitur eksternal pada penelitian lanjutan dapat diuji untuk mengetahui apakah variasi kunjungan yang belum menjelaskan dapat dikurangi tanpa menimbulkan overfitting. Perbandingan dengan Gradient Boosting, XGBoost, CatBoost, dan pendekatan runtun waktu lain juga diperlukan agar keunggulan Random Forest dapat dinilai pada dataset yang sama.

KESIMPULAN DAN SARAN

Kesimpulan

Random Forest berhasil digunakan untuk membangun model prediksi kunjungan harian pasien di Puskesmas Lempuing Kota Bengkulu dengan Mean Absolute Error 7,27 pasien per hari, Root Mean Squared Error 10,25 pasien per hari, dan koefisien determinasi 37,10%. Informasi historis H-7 dan H-1 berperan penting dalam pembentukan prediksi, sementara pola kunjungan menunjukkan variasi menurut hari pelayanan. Integrasi hasil prediksi dengan Standar Beban Kerja sebesar 36 pasien per petugas per hari mampu mengubah keluaran model menjadi rekomendasi kebutuhan 1–3 petugas secara dinamis. Dengan demikian, tujuan penelitian untuk membangun model, mengevaluasi performanya, dan menghasilkan rekomendasi penjadwalan berbasis data dapat dicapai. Keterbatasan utama penelitian adalah periode data yang hanya tiga bulan, belum digunakannya variabel eksternal seperti cuaca, hari libur, dan tren penyakit, serta belum dilakukannya implementasi sistem secara penuh.

Saran

Pengembangan berikutnya disarankan memperluas periode dataset, menambahkan fitur eksternal yang relevan, serta membandingkan Random Forest dengan Gradient Boosting, XGBoost, CatBoost, atau pendekatan runtun waktu lain disertai optimasi hiperparameter. Puskesmas Lempuing juga dapat menguji rekomendasi penjadwalan dinamis ini secara bertahap pada operasi nyata, dengan tetap mempertimbangkan kompetensi petugas, kebutuhan layanan lain, ketentuan kerja, dan evaluasi waktu tunggu pasien. Pada hari dengan proyeksi kunjungan rendah, personel yang tidak dibutuhkan di loket dapat diarahkan ke tugas administrasi atau pelaporan sehingga efisiensi tenaga meningkat tanpa mengurangi kesiapan pelayanan.

DAFTAR PUSTAKA

- Alfarizi, M. R. sirfatullah, Al-farish, M. Z., Taufiqurrahman, M., Ardiansah, G., & Elgar, M. (2023). Penggunaan Python Sebagai Bahasa Pemrograman untuk Machine Learning dan Deep Learning. Karya Ilmiah Mahasiswa Bertauhid (KARIMAH TAUHID), 2(1), 1–6.
- Amaliah, S., Nusrang, M., & Aswi. (2022). Penerapan Metode Random Forest Untuk Klasifikasi Varian Minuman Kopi Di Kedai Kopi Konijiwa Bantaeng. VARIANSI: Journal of Statistics and Its Application on Teaching and Research, 4(2), 121–127.
<https://doi.org/10.35580/variasiunm31>

- Ananda, R., Sumarno, S., & Gunawan, I. (2022). Prediksi Promosi Jabatan Karyawan JNE Pematangsiantar Dengan Algoritma C4.5. *Jurnal Sistem Komputer Dan Informatika (JSON)*, 3(3), 358. <https://doi.org/10.30865/json.v3i3.3763>
- Asaury, A. S., Hamid, A., & Triyono, G. (2025). Prediksi Jumlah Pasien Masuk Rumah Sakit Menggunakan Metode Random Forest. *Jurnal Pendidikan Dan Teknologi Indonesia (JPTI)*, 5(2), 447–459. <https://doi.org/https://doi.org/10.52436/1.jpti.660>
- Ashfania, G. A. M., Prahasto, T., Widodo, A., & Warsokusumo, T. (2023). Penggunaan Algoritma Random Forest untuk Klasifikasi berbasis Kinerja Efisiensi Energi pada Sistem Pembangkit Daya. *ROTASI*, 24(3), 14–21.
- Cubillas, J. J., Ramos, M. I., & Feito, F. R. (2022). Use of Data Mining to Predict the Influx of Patients to Primary Healthcare Centres and Construction of an Expert System. *Applied Sciences*.
- Darmanto, & Barus, E. (2024). Implementasi Metode Random Forest Untuk Memprediksi Penjualan Produk. *Jurnal Tekinkom (Teknik Informasi Dan Komputer)*, 7, 591–600. <https://doi.org/10.37600/tekinkom.v7i2.1510>
- Erkamp, N. S., Dalen, D. H. Van, & Vries, E. De. (2021). Predicting emergency department visits in a large teaching hospital. *International Journal of Emergency Medicine*, 6, 1–11.
- Falah, S., Jaka, A. T., & Nada, N. Q. (2025). Penerapan Algoritma Regresi Linier Dan Random Forest Untuk Memprediksi Penjualan Di Uchi Parfume. *Jurnal Informatika Teknologi Dan Sains (JINTEKS)*, 7(3), 1317–1326.
- Fan, B., Peng, J., Guo, H., Gu, H., Xu, K., & Wu, T. (2022). Accurate Forecasting of Emergency Department Arrivals With Internet Search Index and Machine Learning Models: Model Development and Performance Evaluation. *JMIR Medical Informatics*, 10, 1–17. <https://doi.org/10.2196/34504>
- Febriansyah, F., Dwiyantri, Z. A., Firdaus, D., Informatika, T., Logistik, U., & Nasional, I. T. (2023). Deteksi Serangan Low Rate DDoS pada Jaringan Tradisional menggunakan Algoritma Decision Tree. *Jurnal Cyber Security Dan Forensik Digital*, 6(1), 6–11.
- Febryana, S., Wulandari, I. A., & Prabowo, D. (2024). Rancang Bangun Aplikasi Pengolahan Data Inventory Barang Berbasis Desktop Pada Cv. Hamim Group. *Jurnal Mahasiswa Sistem Informasi (JMSI)*, 6(1), 24–35. <https://doi.org/10.24127/jmsi.v6i1.7553>
- Ihzaniah, L. S., Setiawan, A., & Wijaya, R. W. N. (2023). Perbandingan Kinerja Metode Regresi K-Nearest Neighbor Dan Metode Regresi Linear Berganda Pada Data Boston Housing. *Jambura Journal of Probability and Statistics*, 4(Mei), 17–29.
- Irsabmalfi, R., & Rahmadewi, R. (2025). Deteksi Anatomi Gigi Menggunakan Arsitektur YOLOV8 (You only look once) Untuk Pengenalan Jenis-Jenis Gigi. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 9(3), 4870–4877.
- Koruca, H., Ibrahim, Emek, M. S., & Esra, E. (2023). Development of a new personalized staff-scheduling method with a work-life balance perspective: case of a hospital. *Annal Of Operation Research*, 793–820. <https://doi.org/10.1007/s10479-023-05244-2>

- Naibaho, A. F. A., & Zahra, A. (2023). Prediksi Kelulusan Siswa Sekolah Menengah Pertama Menggunakan Machine Learning. *Jurnal Informatika Dan Teknik Elektro Terapan*, 11(3). <https://doi.org/10.23960/jitet.v11i3.3056>
- Naldy, E. T., & Andri, A. (2021). Penerapan Data Mining Untuk Analisis Daftar Pembelian Konsumen Dengan Menggunakan Algoritma Apriori Pada Transaksi Penjualan Toko Bangunan MDN. *Jurnal Nasional Ilmu Komputer*, 2(2), 89–101. <https://doi.org/10.47747/jurnalnik.v2i2.525>
- Noija, H. C., Wemaf, P. A., Nurdianty, O. A., Sohilit, W., Haumahu, S., Yusuf, H., Tomagola, K., Wa Rindi La Ana, & Salaiswa, T. S. (2023). Perancangan Sistem Informasi Akuntansi Untuk Analisis Siklus Pendapatan Pada Orantata Celullar Menggunakan DFD Dan Flowchart. *Jurnal Bisnis Dan Manajemen (JURBISMAN)*, 1(2), 577–592. <https://doi.org/10.61930/jurbisman.v1i2.188>
- Nugroho, B. I., Fatkhurrozak, Andriani, W., & Gunawan, G. (2024). Prediksi Jumlah Kunjungan Pasien Rawat Jalan Menggunakan Metode Resgresi Linier Sederhana. *PROSISKO: Jurnal Pengembangan Riset Dan Observasi Sistem Komputer*, 11(2), 284–293.
- Octarina, S., Puspita, F. M., Yuliza, E., & Indrawati. (2025). Pendampingan Penggunaan Google Colab Pada Pembelajaran Python Dan Machine Learning Bagi Dosen Matematika Di Palembang. *Jurnal Terpadu*, 6(1), 56–66.
- Pranajaya, A. E., & Susanto, E. R. (2025). Optimasi Random Forest untuk Prediksi Penyakit Jantung Menggunakan SMOTEENN dan Grid Search. *Jurnal Pendidikan Dan Teknologi Indonesia (JPTI)*, 5(7), 1965–1979. <https://doi.org/https://doi.org/10.52436/1.jpti.855>
- Rahmat, A., Hardi, H., Syam, F. A., Zamzami, Z., Febriadi, B., & Windarto, A. P. (2021). Utilization of the field of data mining in mapping the area of the Human Development Index (HDI) in Indonesia. *Journal of Physics: Conference Series*, 1783(1). <https://doi.org/10.1088/1742-6596/1783/1/012035>
- Sartika, D., & Jumadi, J. (2023). Implementasi Data Mining Untuk Pengelompokkan Wilayah Pelanggaran Lalu Lintas Menggunakan Metode K-Means Pada Polres Bengkulu. *Journal of Science and Social Research*, 4307(2), 345–349. <http://jurnal.goretanpena.com/index.php/JSSR>
- Selya, A., Anshutz, D., Griesse, E., Weber, T. L., Hsu, B., & Ward, C. (2021). Predicting unplanned medical visits among patients with diabetes : translation from machine learning to clinical implementation. *BMC Medical Informatics And Diecision Making*, 1–11.
- Sultan, M., Darnila, E., & Afrillia, Y. (2025). Prediksi Potensi Wisata Menggunakan Algoritma Random Forest Di Kabupaten Gayo Lues. *Rabit : Jurnal Teknologi Dan Sistem Informasi Univrab*, 10(2), 741–751. <https://doi.org/10.36341/rabit.v10i2.6365>
- Surbakti, N. M., Angelyca, Talia, A., Perangin-Angin, C. B., Nainggolan, D. O., Friskauly, N. D., & Tumorang, S. R. B. (2024). Penggunaan Bahasa Pemrograman Python dalam Pembelajaran Kalkulus Fungsi Dua Variabel. *Algoritma: Jurnal Matematika, Ilmu Pengetahuan Alam, Kebumian Dan Angkasa*, 2(3), 98–107.

- Tuzzakiyah, E., Kartika, R. C., Ayu, D. P., Fitriyah, D., & Puspita, S. D. (2022). Analisis Kebutuhan Tenaga Rekam Medis dengan Metode Workload Indicator Staffing Need (WISN). *Jurnal Rekam Medik & Manajemen Informasi Kesehatan*, 1(2), 73–79.
- Wardhana, R. G., Wang, G., & Sibuea, F. (2023). Penerapan Machine Learning Dalam Prediksi Tingkat Kasus Penyakit Di Indonesia. *Journal of Information System Management (JOISM)*, 5(1).
- Widodo, A. P., & Setiawan, A. (2025). Systematic Literature Review : Peran Machine Learning dalam Manajemen Sumber Daya Manusia. *Optimal: Jurnal Ekonomi Dan Manajemen*, 5, 650–665. <https://doi.org/https://doi.org/10.55606/optimal.v5i4.8089>
- Wiguna, D. P., Juliana, L., & Ginting, S. E. (2025). Aplikasi Google Colab Berbasis Python dalam Menerapkan Teori Pohon dengan Algoritma Random Forest Classifier. *Scripta Technica : Journal of Engineering and Applied Technology*, 1(2), 261–270.
- Wulandari, G., Sasmita, S., & Rahmadayanti, F. (2024). Implementasi Data Mining Menggunakan Algoritma Apriori Dalam Menentukan Persediaan Obat. *Jurnal Teknik Informasi Dan Komputer (Tekinkom)*, 7(1), 144. <https://doi.org/10.37600/tekinkom.v7i1.1270>