

Prediksi Lama Rawat Inap Pasien POSTPARTUM Berdasarkan Faktor Resiko Klinis dan Demografis Menggunakan Algoritma Random Forest

Maya Nurul Nabila*, Iman Paryudi²

Jurusan Teknik Informatika, Fakultas Teknik, Universitas Pancasila, Jakarta

Abstrak. Lama rawat inap pasien postpartum merupakan indikator penting dalam pelayanan rumah sakit, namun penentuannya masih sering dilakukan secara manual dan subjektif. Penelitian ini bertujuan mengembangkan model prediksi durasi rawat inap berbasis *Random Forest Regressor* yang diintegrasikan dalam aplikasi website untuk memberikan estimasi otomatis berdasarkan data klinis dan demografis. Model dilatih menggunakan sepuluh atribut (usia, gravida, parity, abortus, jenis persalinan, tekanan darah, hemoglobin, komplikasi kehamilan, riwayat penyakit, dan metode pembayaran) dan menghasilkan performa yang baik pada data uji dengan nilai RMSE 0,39, MAE 0,16, MAPE 0,05, dan R^2 0,72 dari rata-rata aktual 2,64 hari. Aplikasi yang dibangun dengan Streamlit mampu memberikan prediksi *real-time* secara konsisten, sehingga berpotensi mendukung keputusan medis yang lebih cepat dan tepat. Penelitian selanjutnya dapat mengembangkan model dengan data yang lebih besar, menambahkan variabel klinis, mencoba algoritma lain, serta mengintegrasikan sistem dengan platform informasi rumah sakit.

Kata kunci— *Postpartum*, Rawat Inap, *Random Forest Regressor*, Prediksi

1. PENDAHULUAN

Pelayanan kesehatan yang berkualitas merupakan indikator penting dalam upaya meningkatkan derajat kesehatan masyarakat. Salah satu aspek yang digunakan untuk menilai kualitas pelayanan rumah sakit adalah lama rawat inap atau *length of stay* (LOS), yaitu durasi perawatan pasien sejak masuk hingga diperbolehkan pulang. LOS menjadi tolak ukur penting karena mencerminkan efektivitas manajemen pelayanan rumah sakit. Lama rawat inap yang berlebihan dapat meningkatkan biaya, memperbesar risiko infeksi nosokomial, serta membatasi ketersediaan tempat tidur bagi pasien lain yang membutuhkan[1].

Masa *postpartum* atau nifas merupakan periode kritis dalam kehidupan seorang perempuan, ditandai dengan perubahan fisik maupun emosional yang signifikan[2]. Pemantauan intensif pada periode ini diperlukan untuk menjamin kesehatan ibu dan bayi serta mencegah komplikasi. Durasi rawat inap pasien postpartum menjadi salah satu aspek penting yang berpengaruh pada penggunaan sumber daya, biaya perawatan, dan kualitas layanan kesehatan[3].

Di rumah sakit, termasuk RSUD Balaraja, penentuan lama rawat inap postpartum umumnya didasarkan pada pengalaman tenaga medis dan pedoman klinis yang berlaku. Namun, setiap pasien memiliki kondisi yang berbeda, seperti riwayat kehamilan, komplikasi persalinan, tekanan darah, usia, serta faktor sosial yang memengaruhi kebutuhan perawatan[4]. Oleh karena itu, diperlukan pendekatan berbasis data untuk menghasilkan keputusan yang lebih tepat dan sesuai dengan kondisi individu.

Perkembangan teknologi digital memungkinkan penggunaan *machine learning* dalam mendukung pengambilan keputusan klinis secara lebih akurat. *Machine learning* dapat mengolah data medis yang kompleks untuk memberikan wawasan prediktif secara otomatis[5]. Salah satu algoritma yang banyak

* Corresponding author: mayanabila82@gmail.com

digunakan adalah Random Forest, yang mampu menangani data dengan banyak variabel, menghasilkan prediksi dengan tingkat kesalahan rendah, serta relatif mudah diinterpretasikan[6]. Penelitian terdahulu menunjukkan bahwa Random Forest memiliki performa unggul dibandingkan algoritma lain, misalnya pada prediksi lama rawat inap pasien Covid-19, di mana Random Forest memperoleh akurasi tertinggi sebesar 73,3% dibandingkan K-Nearest Neighbor, Support Vector Machine, dan Gradient Boosting[7].

Selain faktor klinis, sejumlah penelitian juga menekankan pentingnya faktor demografis dalam pemodelan berbasis *machine learning*. Studi [8] menunjukkan bahwa variabel demografi, seperti usia, jenis kelamin, dan latar belakang rasial, dapat dimanfaatkan secara signifikan untuk memodelkan kepribadian individu. Temuan ini mendukung bahwa faktor demografis memiliki kontribusi besar dalam meningkatkan kinerja prediksi di berbagai domain, sehingga relevan dengan penelitian ini yang memanfaatkan faktor risiko klinis dan demografis dalam memprediksi lama rawat inap pasien *postpartum*. Penelitian lain [9] bahkan mengungkapkan bahwa informasi sederhana seperti usia dan jenis kelamin sudah cukup untuk memperkirakan karakteristik individu tanpa memerlukan variabel tambahan. Hal ini memperlihatkan bahwa data demografis memiliki hubungan yang kuat dengan hasil prediksi, sehingga layak digabungkan bersama faktor klinis dalam upaya memperkirakan lama rawat inap pasien *postpartum*.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk mengembangkan model prediksi lama rawat inap pasien *postpartum* menggunakan algoritma Random Forest dengan mempertimbangkan faktor klinis dan demografis. Model yang dihasilkan diharapkan mampu mendukung pengambilan keputusan medis, meningkatkan efisiensi manajemen rumah sakit, serta memastikan pasien memperoleh perawatan sesuai dengan kebutuhan klinisnya.

2. METODE PENELITIAN

Pasien *postpartum* membutuhkan perhatian khusus selama masa pemulihan, dan durasi rawat inap menjadi indikator penting yang memengaruhi kualitas layanan serta efisiensi pengelolaan rumah sakit. Faktor klinis maupun demografis, seperti usia, riwayat kehamilan, jenis persalinan, komplikasi, tekanan darah, kadar hemoglobin, dan metode pembayaran, turut menentukan lamanya perawatan. Untuk mengatasi kompleksitas tersebut, penelitian ini membangun model prediksi lama rawat inap dengan algoritma Random Forest Regression yang mampu menangani data medis bervariasi, menghasilkan prediksi konsisten, serta mengidentifikasi faktor paling berpengaruh. Model ini diharapkan dapat mendukung pengambilan keputusan medis dan pengelolaan sumber daya di RSUD Balaraja secara lebih efektif.

2.1. Tahapan Penelitian

Tahapan penelitian ini terdiri dari beberapa tahapan utama, yaitu:

1. Pengumpulan data

Data yang digunakan merupakan data internal RSUD Balaraja periode Januari–Desember 2024, mencakup faktor klinis dan demografis pasien *postpartum* yang menjalani rawat inap. Data diperoleh dari sistem rekam medis rumah sakit, kemudian diekspor dalam format *Comma Separated Values* (.CSV) agar dapat diolah menggunakan Python dan pustaka *machine learning*.

2. Pra-pemrosesan data

Data mentah melalui tahapan pembersihan (*data cleaning*), penanganan nilai hilang (*missing values*), serta transformasi data ke bentuk numerik agar sesuai untuk pemodelan. Selanjutnya dilakukan pembagian data menjadi data latih dan data uji dengan perbandingan 80:20.

3. Pemodelan dengan Random Forest

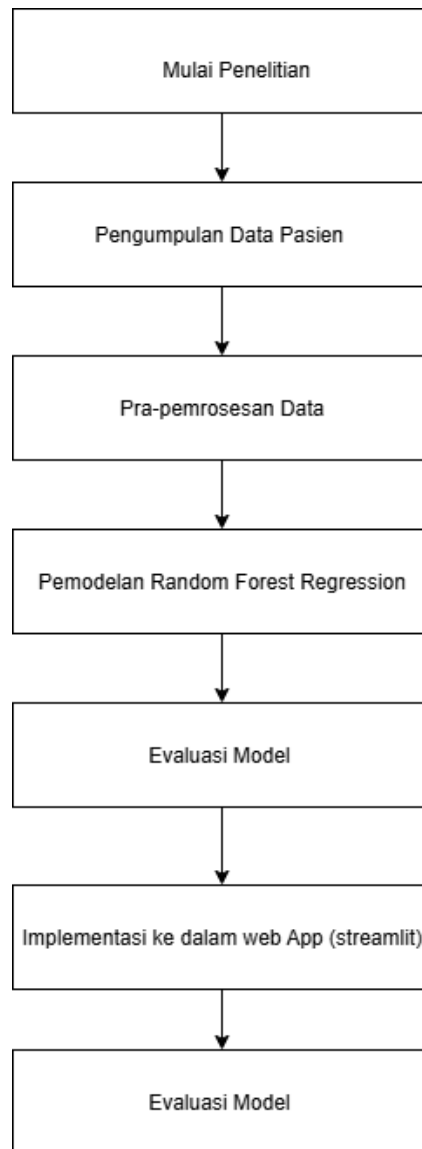
Algoritma Random Forest Regression dipilih karena mampu menangani data kompleks dengan variabel yang bervariasi [10]. Model dibangun dengan jumlah *estimators* tertentu, kemudian dilatih menggunakan data latih. Selain menghasilkan prediksi lama rawat inap, model ini juga digunakan untuk mengidentifikasi faktor-faktor yang paling berpengaruh terhadap durasi perawatan pasien.

4. Evaluasi Model

Evaluasi dilakukan menggunakan metrik *Root Mean Square Error* (RMSE), *Mean Absolute Error* (MAE), *Mean Absolute Percentage Error* (MAPE), dan *Koefisien Determinasi* (R^2) [11]. Visualisasi berupa *decision tree* dari beberapa pohon dalam Random Forest serta grafik evaluasi digunakan untuk memperkuat interpretasi hasil.

5. Perancangan dan Implementasi Aplikasi

Model yang telah dibangun diimplementasikan ke dalam aplikasi berbasis web menggunakan framework *Streamlit*. Arsitektur sistem terdiri dari antarmuka pengguna, modul pemrosesan data, dan modul prediksi berbasis Random Forest. Aplikasi ini dirancang untuk menerima input faktor risiko pasien postpartum, kemudian memberikan output berupa estimasi lama rawat inap secara *real-time*



Gambar 1 *flow chart* penelitian

2.2. Data

Penelitian ini menggunakan data internal RSUD Balaraja yang mencakup informasi klinis dan demografis pasien *postpartum* yang dirawat inap pada periode Januari–Desember 2024. Data diperoleh dari sistem rekam medis rumah sakit, kemudian disusun dalam bentuk tabel dan diekspor ke format *Comma Separated Values* (.CSV) untuk selanjutnya diolah menggunakan bahasa pemrograman Python dengan pustaka *machine learning*.

3. HASIL DAN PEMBAHASAN

3.1 Analisis situasional

Lama rawat inap (*length of stay*/LOS) merupakan salah satu indikator utama dalam menilai kualitas pelayanan rumah sakit. LOS yang terlalu panjang dapat meningkatkan risiko komplikasi, biaya perawatan, dan kebutuhan sumber daya tambahan, sedangkan LOS yang terlalu pendek dapat meningkatkan peluang pasien kembali ke rumah sakit akibat perawatan yang belum optimal. Kondisi ini sejalan dengan penelitian [1] yang

menegaskan bahwa LOS memiliki hubungan erat dengan efektivitas manajemen pelayanan kesehatan serta ketersediaan tempat tidur.

Pada masa postpartum, kebutuhan perawatan setiap pasien sangat beragam. Faktor seperti jenis persalinan, komplikasi obstetri, tekanan darah, usia, dan kadar hemoglobin memiliki peran signifikan dalam memengaruhi kecepatan pemulihan ibu setelah melahirkan [2][3]. Di RSUD Balaraja, penentuan lama rawat inap umumnya masih mengandalkan penilaian klinis tenaga medis. Namun, heterogenitas kondisi pasien sering kali menjadikan proses penilaian tersebut bersifat subjektif dan berpotensi menimbulkan ketidakkonsistenan.

Seiring perkembangan teknologi, pendekatan machine learning menawarkan analisis berbasis data yang lebih objektif dan prediktif. Machine learning terbukti mampu mengolah variabel klinis yang kompleks serta menghasilkan prediksi kesehatan dengan akurasi tinggi [5]. Salah satu algoritma yang banyak digunakan adalah Random Forest, yang dikenal memiliki performa prediktif stabil serta bekerja baik pada data multivariat [6]. Beberapa penelitian menunjukkan bahwa Random Forest bahkan lebih unggul dibandingkan algoritma lain seperti KNN, SVM, dan Gradient Boosting dalam memprediksi LOS pada berbagai kasus medis termasuk pasien Covid-19 [7].

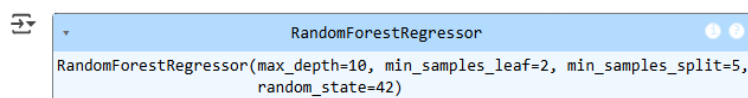
Selain faktor klinis, variabel demografis seperti usia dan jenis kelamin juga diketahui memberikan kontribusi penting dalam memodelkan karakteristik dan hasil kesehatan [8][9]. Oleh karena itu, penggabungan variabel klinis dan demografis dalam model prediksi menjadi langkah strategis untuk mendapatkan estimasi LOS yang lebih akurat.

Berdasarkan kondisi tersebut, pembangunan model prediksi LOS postpartum menggunakan algoritma Random Forest Regression menjadi relevan bagi RSUD Balaraja untuk meningkatkan efisiensi manajemen perawatan, mendukung pengambilan keputusan klinis, serta menghasilkan estimasi LOS yang lebih konsisten dan berbasis data.

3.2. Hasil Pemodelan

Model prediksi lama raat inap pasein *postpartum* dibangun menggunakan algoritma Random Forest Regressor dengan konfigurasi parameter terbaik hasil *tuning grid search* dan validasi silang. Parameter akhir yang digunakan adalah:

- Max_depth = 10
- Min_samples_leaf = 2
- Min_samples_split = 5
- Random_state = 42



Gambar 2 Pemodelan dengan Random Forest Regression

Kombinasi parameter ini dipilih karena memberikan performa paling optimal sekaligus menjaga keseimbangan antara kompleksitas model dan kemampuan generalisasi. Evaluasi model dilakukan menggunakan skema **5-fold cross validation** dan empat metrik regresi, yaitu RMSE, MAE, MAPE, dan koefisien determinasi (R^2).

Tabel 1 Hasil Evaluasi Model

RMSE	MAE	MAPE	R^2
0.39	0.16	0.05	0.72

Nilai **R² sebesar 0,72** menunjukkan bahwa model mampu menjelaskan 72% variasi data target (lama rawat inap), sementara nilai **RMSE (0,39)** dan **MAE (0,16)** menandakan kesalahan prediksi relatif kecil. Nilai **MAPE sebesar 5%** juga memperkuat bahwa model memiliki tingkat kesalahan yang rendah dalam memprediksi durasi rawat inap.

Rumus masing – masing metrik evaluasi regresi yang digunakan adalah sebagai berikut:

- a. Root Mean Square Error (RMSE)

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_i)^2}$$

- b. Mean Absolute Error (MAE)

$$MAE = \frac{1}{n} \sum_{i=1}^n |x_i - \bar{x}_i|$$

- c. Mean Absolute Percentage Error (MAPE)

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{x_i - \bar{x}_i}{\bar{x}_i} \right|$$

- d. Koefisien Deterinasi (R²)

$$R^2 = 1 - \frac{\sum_{i=1}^n (x_i - \hat{x}_i)^2}{\sum_{i=1}^n (x_i - \bar{x}_i)^2}$$

Sebagai ilustrasi, sebagian hasil perbandingan nilai aktual dan prediksi ditampilkan pada tabel 2. Hasil ini memperlihatkan bahwa prediksi model mendekati nilai sebenarnya dengan selisih yang relative kecil.

Tabel 2 Perbandingan Nilai Prediksi dan data Asli

No	Nilai Sebenarnya	Hasil prediksi
1	2	2.000000
2	2	2.000000
3	5	3.207940
4	2	2.000000
5	3	3.018667

Visualisasi perbandingan nilai aktual dan prediksi ditunjukkan pada Tabel 2. Grafik ini menunjukkan bahwa model mampu mengikuti pola distribusi data dengan baik, meskipun masih terdapat deviasi pada nilai ekstrem.

Hasil penelitian menunjukkan bahwa Random Forest Regression mampu memberikan prediksi lama rawat inap pasien postpartum secara cukup akurat. Model dapat mengidentifikasi faktor-faktor utama yang berpengaruh terhadap durasi rawat inap, sehingga dapat berfungsi sebagai alat bantu pengambilan keputusan klinis di RSUD Balaraja.

Dibandingkan pendekatan tradisional yang mengandalkan intuisi dan pengalaman tenaga medis, pendekatan berbasis data menghasilkan prediksi yang lebih objektif dan konsisten. Hal ini sejalan dengan penelitian [5] yang menegaskan bahwa machine learning mampu menangani data medis kompleks dengan lebih akurat dibandingkan metode manual.

Selain itu, hasil penelitian memperkuat temuan dalam literatur bahwa faktor demografis seperti usia juga berperan penting dalam membentuk hasil kesehatan [8][9]. Penggabungan variabel klinis dan demografis terbukti meningkatkan kemampuan model dalam memprediksi LOS.

Secara keseluruhan, model prediksi yang dibangun berhasil menunjukkan bahwa algoritma Random Forest merupakan metode yang stabil, andal, serta memiliki performa unggul dalam memprediksi outcome kesehatan [6][7][10]. Dengan akurasi yang cukup tinggi serta kemampuan mengidentifikasi faktor penting, model ini berpotensi mendukung pengambilan keputusan berbasis data dan meningkatkan efisiensi manajemen layanan kesehatan, khususnya di RSUD Balaraja.

4. KESIMPULAN

Penelitian ini menunjukkan bahwa algoritma Random Forest Regressor dapat digunakan secara efektif untuk memprediksi lama rawat inap pasien postpartum di RSUD Balaraja. Model yang dibangun memanfaatkan sepuluh atribut klinis dan demografis, meliputi usia, riwayat kehamilan, jenis persalinan, tekanan darah, hemoglobin, komplikasi kehamilan, riwayat penyakit, dan metode pembayaran. Hasil evaluasi menghasilkan nilai RMSE = 0,39, MAE = 0,16, MAPE = 0,05, dan $R^2 = 0,72$, yang mengindikasikan tingkat kesalahan prediksi yang rendah serta kemampuan model dalam menjelaskan 72% variasi data target.

Selain memberikan estimasi durasi rawat inap dengan cukup akurat, model juga mengidentifikasi faktor-faktor klinis yang paling berpengaruh, khususnya jenis persalinan, riwayat penyakit, tekanan darah, dan usia. Dengan demikian, model ini berpotensi menjadi alat bantu dalam pengambilan keputusan medis berbasis data serta mendukung efisiensi manajemen layanan kesehatan, khususnya pada pasien *postpartum*.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada semua pihak yang telah memberikan dukungan dan bantuan selama proses penelitian ini. Secara khusus, penulis menyampaikan apresiasi kepada dosen pembimbing atas arahan, masukan, dan bimbingannya sehingga penelitian ini dapat terselesaikan dengan baik. Diharapkan hasil penelitian ini dapat memberikan kontribusi positif bagi pengembangan ilmu pengetahuan di bidang yang relevan.

DAFTAR PUSTAKA

- [1] N. L. Revalin, “PENGARUH PERBAIKAN MUTU TERHADAP PENURUNAN LAMA RAWAT INAP (LENGTH OF STAY) : LITERATURE REVIEW,” vol. 5, pp. 13094–13103, 2024.
- [2] R. P. Wulandari and M. Mufdlilah, “Faktor demografi dan obstetrik dalam mempengaruhi kualitas hidup postpartum,” *J. Kebidanan*, vol. 9, no. 2, p. 129, 2020, doi: 10.26714/jk.9.2.2020.129-142.
- [3] R. A. Rusadi, Musdalifah, and Y. Iskandar, “Lama Waktu Rawat Inap Pasien Operasi Bedah Caesar,” *J. Bidan*, vol. 1, no. 1, pp. 24–28, 2023, [Online]. Available: <https://journal.umbogorraya.ac.id/index.php/JB/article/view/298>
- [4] B. A. Pratama, “Hubungan Usia Kehamilan dan Jenis Persalinan dengan Lama Rawat Inap Kasus Low Birth Weight di RS PKU Muhammadiyah Surakarta Tahun 2022,” *Indones. J. Med. Sci.*, vol. 11, no. 1, pp. 63–70, 2024, doi: 10.55181/ijms.v11i1.474.
- [5] R. G. Wardhana, G. Wang, and F. Sibuea, “Penerapan Machine Learning Dalam Prediksi Tingkat Kasus Penyakit Di Indonesia,” *J. Inf. Syst. Manag.*, vol. 5, no. 1, pp. 40–45, 2023, doi: 10.24076/joism.2023v5i1.1136.
- [6] R. P. Munggaran and M. Nurmalasari, “Predicting Outpatient Service Waiting Times with Random Forest Algorithm Prediksi Waktu Tunggu Pelayanan Pasien Rawat Jalan dengan Algoritma Random Forest,” vol. 5, no. January, pp. 35–40, 2025.
- [7] P. Syahputra, “Prediksi Lama Rawat Pasien Covid-19 Berbasis Machine Learning,” *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 9, no. 4, pp. 3374–3382, 2022, doi: 10.35957/jatisi.v9i4.2883.
- [8] I. Paryudi, E. Winarko, S. Priyanta, and S. R. C. Nursari, “Modeling of Personality Traits based on Demographic Data on Multi-Races Samples of Ages from 13 to 50 Years Old: Investigating the Effect of Race on Model,” *Proc. - 2020 6th Int. Conf. Sci. Technol. ICST 2020*, pp. 2–7, 2020, doi: 10.1109/ICST50505.2020.9732792.
- [9] I. Paryudi, A. Ashari, and A. M. Tjoa, “Personality Estimation using Demographic Data in a

- Personality-based Recommender System: A Proposal,” *ACM Int. Conf. Proceeding Ser.*, 2019, doi: 10.1145/3366030.3366098.
- [10] R. Supriyadi, W. Gata, N. Maulidah, and A. Fauzi, “Penerapan Algoritma Random Forest Untuk Menentukan Kualitas Anggur Merah,” *E-Bisnis J. Ilm. Ekon. dan Bisnis*, vol. 13, no. 2, pp. 67–75, 2020, doi: 10.51903/e-bisnis.v13i2.247.
- [11] L. Benedict and U. M. Nusantara, “Prediksi Tingkat Kematian Covid-19 di Indonesia dengan menggunakan Metode Linear Regression,” pp. 5–12, 2022, [Online]. Available: <https://kc.umn.ac.id/id/eprint/22407>