



**UJMAS**

UMBARA JOURNAL OF MATHEMATICS, ACTUARIAL SCIENCE AND STATISTICS  
<https://journal.umbogorraya.ac.id/index.php/ujmas>

## Analisis Risiko Mortalitas Menggunakan Algoritma Random Forest

Florenta<sup>1</sup>, Suwarjono<sup>2</sup>, Dyah Prita Anggraini<sup>3</sup>, Elsa Elsa Megamarc<sup>4</sup>

<sup>1,3,4</sup> Program Studi Sains Aktuaria, Fakultas Kesehatan dan Sains

<sup>2</sup> Program Studi Ilmu Komputer, Fakultas Kesehatan dan Sains

Universitas Muhammadiyah Bogor Raya

Jln. Raya Leuwiliang No. 106 Kec. Leuwiliang Kabupaten Bogor

Volume 1 Nomor 1  
Desember 2024 : 35 - 42

### Article History

Submission: 09-12-2024

Revised: 09-12-2024

Accepted: 15-12-2024

Published: 20-12-2024

### Kata Kunci:

Prediksi Mortalitas; Random Forest;  
Pasien Rawat Inap; Faktor Risiko  
Klinis; Sistem Peringatan Dini Klinis.

### Keywords:

Decision Support System; Insurance  
Premium; Data Mining; Random Forest;  
Risk Prediction.

### Korespondensi:

(Florenta)

(Telp.-)

([florentaasukendar@gmail.com](mailto:florentaasukendar@gmail.com))

**Abstrak:** Penelitian ini bertujuan untuk mengembangkan model prediksi risiko mortalitas pasien rawat inap berbasis algoritma Random Forest yang akurat, stabil, dan reliabel sebagai dasar sistem peringatan dini klinis. Data penelitian menggunakan data rekam medis pasien rawat inap yang telah melalui proses pembersihan, normalisasi, serta penyeimbangan kelas. Model dilatih menggunakan skema 5-fold cross validation untuk menjamin kestabilan performa prediksi. Hasil evaluasi menunjukkan bahwa model Random Forest memiliki kinerja yang sangat baik dengan nilai akurasi sebesar 87,4%, sensitivitas 85,1%, spesifisitas 88,6%, dan AUC 0,91, sehingga memenuhi kriteria sebagai excellent classifier. Analisis feature importance mengidentifikasi lima faktor klinis utama yang paling berpengaruh terhadap risiko mortalitas pasien rawat inap, yaitu usia lanjut, saturasi oksigen, tekanan darah sistolik, kadar hemoglobin, dan riwayat komorbid. Model mampu mengklasifikasikan pasien ke dalam tiga kategori risiko, yaitu risiko rendah (<10%), sedang (10–30%), dan tinggi (>30%), yang efektif membedakan tingkat probabilitas kematian aktual. Model prediksi yang dikembangkan kemudian dirancang sebagai dasar sistem peringatan dini klinis yang mampu mengidentifikasi pasien berisiko tinggi secara lebih cepat dan akurat, sehingga berpotensi meningkatkan keselamatan pasien serta mutu pelayanan rumah sakit. Hasil penelitian ini menunjukkan bahwa Random Forest layak digunakan sebagai komponen utama dalam pengembangan sistem peringatan dini klinis berbasis kecerdasan buatan.

**Abstract:** This study aims to develop a Random Forest-based prediction model for inpatient mortality risk that is accurate, stable, and reliable as the basis for a clinical early warning system. The study uses inpatient electronic medical record data that have undergone data cleaning, normalization, and class balancing. The model was trained using a 5-fold cross-validation scheme to ensure stable predictive performance. The evaluation results show that the



UJMAS: Umbara Journal Of Mathematics, Actuarial Science And Statistics is licensed under a Creative Commons Attribution-Share Alike 4.0 International License. Copyright © 2025 Prodi Sains Aktuaria Universitas Muhammadiyah Bogor Raya, Indonesia. All Rights Reserved

---

*Random Forest model achieved excellent performance with an accuracy of 87.4%, sensitivity of 85.1%, specificity of 88.6%, and an AUC of 0.91, thus meeting the criteria of an excellent classifier. Feature importance analysis identified five main clinical factors that most strongly influence inpatient mortality risk, namely advanced age, oxygen saturation, systolic blood pressure, hemoglobin level, and comorbidity history. The model classifies patients into three risk categories: low risk (<10%), moderate risk (10–30%), and high risk (>30%), which effectively differentiate actual mortality probabilities. The developed prediction model is designed as the core component of a clinical early warning system to enable earlier and more accurate identification of high-risk patients, thereby potentially improving patient safety and the quality of hospital care. This study demonstrates that Random Forest is a suitable and effective approach for developing AI-based clinical early warning systems.*

---

## PENDAHULUAN

Mortalitas masih menjadi indikator utama mutu layanan kesehatan dan keberhasilan sistem kesehatan di berbagai negara, termasuk Indonesia. Peningkatan kasus penyakit tidak menular, komplikasi akut, serta keterbatasan tenaga dan fasilitas kesehatan menyebabkan tingginya beban layanan rawat inap dan gawat darurat. Data nasional menunjukkan bahwa sebagian besar kematian rumah sakit terjadi akibat keterlambatan penanganan pasien berisiko tinggi, terutama pada fase awal perawatan [1]. Hal ini menunjukkan urgensi pengembangan sistem prediksi dini yang mampu mengidentifikasi pasien dengan risiko mortalitas tinggi secara cepat dan akurat.

Model prediksi mortalitas yang saat ini banyak digunakan masih didominasi oleh pendekatan statistik konvensional seperti regresi logistik dan model Cox proportional hazard [2]. Meskipun mudah diinterpretasikan, model ini

memiliki keterbatasan dalam menangkap hubungan nonlinier dan interaksi kompleks antarvariabel klinis [3]. Padahal, data klinis modern umumnya bersifat berdimensi tinggi, heterogen, dan mengandung pola yang kompleks [4]. Oleh karena itu, pendekatan berbasis *machine learning* menjadi semakin relevan untuk meningkatkan akurasi prediksi risiko mortalitas.

Algoritma Random Forest merupakan salah satu metode *ensemble learning* yang terbukti memiliki performa tinggi dalam klasifikasi data medis, khususnya untuk data tidak seimbang dan nonlinier [5]. Beberapa studi internasional melaporkan bahwa Random Forest mampu memberikan akurasi dan Area Under Curve (AUC) yang lebih tinggi dibandingkan metode statistik klasik dalam prediksi mortalitas pasien rawat inap dan ICU [6,7]. Namun, pemanfaatan Random Forest dalam konteks rumah sakit di

Indonesia masih sangat terbatas, khususnya yang terintegrasi dengan analisis *feature importance* untuk mendukung pengambilan keputusan klinis.

Berdasarkan kondisi tersebut, rumusan permasalahan dalam penelitian ini adalah: (1) bagaimana karakteristik model prediksi risiko mortalitas pasien rawat inap berbasis Random Forest yang akurat dan reliabel; (2) faktor klinis apa saja yang paling berpengaruh terhadap risiko mortalitas pasien; dan (3) bagaimana merancang model prediksi yang dapat dijadikan dasar sistem peringatan dini klinis.

#### METODE

Penelitian ini menggunakan pendekatan kuantitatif dengan desain *predictive analytics research* untuk mengembangkan model prediksi risiko mortalitas pasien rawat inap berbasis algoritma Random Forest. Penelitian dilaksanakan melalui pengolahan data sekunder rekam medis rumah sakit mitra dan menghasilkan model prediksi yang siap diintegrasikan ke sistem informasi rumah sakit sebagai sistem peringatan dini klinis berbasis kecerdasan buatan.

Penelitian ini diawali dengan kegiatan pengumpulan data sekunder berupa rekam medis pasien rawat inap dari rumah sakit mitra yang telah melalui proses perizinan dan persetujuan etik. Data yang dikumpulkan meliputi karakteristik demografis pasien, hasil pemeriksaan klinis dan laboratorium, diagnosis penyakit, riwayat tindakan medis, serta status keluaran pasien (hidup atau meninggal). Data tersebut kemudian diseleksi berdasarkan kriteria inklusi dan eksklusi untuk memastikan kesesuaian dengan tujuan penelitian.

Tahap selanjutnya adalah praproses data yang meliputi pembersihan data dari duplikasi dan kesalahan pencatatan, penanganan data hilang menggunakan metode imputasi yang sesuai, serta normalisasi data numerik. Mengingat data mortalitas umumnya tidak seimbang antara kelas hidup dan meninggal, dilakukan pula penyeimbangan kelas menggunakan teknik *resampling* agar model dapat belajar secara optimal.

Setelah data siap diolah, dilakukan seleksi variabel klinis utama yang relevan terhadap risiko mortalitas

berdasarkan kajian literatur dan analisis statistik awal. Variabel-variabel ini menjadi dasar dalam pembangunan model prediksi menggunakan algoritma Random Forest. Model dilatih menggunakan teknik *k-fold cross validation* untuk memastikan stabilitas dan generalisasi model terhadap data baru.

Model yang dihasilkan kemudian dievaluasi berdasarkan indikator akurasi, sensitivitas, spesifisitas, dan Area Under Curve (AUC). Hasil evaluasi digunakan untuk melakukan optimasi parameter model sehingga diperoleh performa prediksi terbaik. Selanjutnya dilakukan analisis *feature importance* untuk mengidentifikasi faktor-faktor klinis yang paling berpengaruh terhadap risiko mortalitas, sehingga hasil penelitian tidak hanya akurat tetapi juga dapat diinterpretasikan secara klinis.

## HASIL & PEMBAHASAN

### Hasil

Hasil pelaksanaan penelitian berupa Model Prediksi Risiko Mortalitas Berbasis Random Forest. Hasil pengembangan Model Random Forest berhasil dikembangkan menggunakan

data rekam medis pasien rawat inap yang telah melalui proses pembersihan, normalisasi, serta penyeimbangan kelas. Model dilatih menggunakan skema 5-fold cross validation untuk menjamin kestabilan performa prediksi. Hasil evaluasi menunjukkan bahwa model Random Forest memiliki kinerja prediksi yang sangat baik dalam mengklasifikasikan risiko mortalitas pasien.

Tabel 1.1 Hasil Evaluasi Pengembangan Model

Random Forest	
Indikator	Hasil
Akurasi	87,4%
Sensitivitas	85,1%
Spesifisitas	88,6%
AUC	0,91

Capaian ini telah **melampaui indikator luaran penelitian** (akurasi  $\geq 80\%$  dan AUC  $\geq 0,80$ ), sehingga model dinyatakan layak sebagai sistem prediksi dini klinis.

### Hasil Identifikasi Faktor Risiko Mortalitas

Analisis *feature importance* Random Forest menghasilkan lima faktor klinis paling dominan yang berpengaruh terhadap risiko mortalitas pasien rawat inap:

Tabel 1.2 Hasil Identifikasi Faktor Risiko Mortalitas

Peringkat	Faktor Risiko Klinis
1	Usia lanjut
2	Saturasi oksigen
3	Tekanan darah sistolik
4	Kadar hemoglobin
5	Riwayat komorbid

Hasil ini menunjukkan bahwa parameter vital dan status klinis awal pasien merupakan indikator paling kuat terhadap risiko kematian.

### Klasifikasi Risiko Pasien

Model mengklasifikasikan pasien ke dalam tiga kategori risiko:

Tabel 1.3 Klasifikasi Risiko Pasien

Kategori Risiko	Probabilitas Kematian
Rendah	< 10%
Sedang	10% - 30%
Tinggi	> 30%

Distribusi hasil menunjukkan bahwa pasien risiko tinggi memiliki proporsi mortalitas aktual paling besar, sehingga model efektif digunakan sebagai alat peringatan dini klinis.

### Pembahasan

Karakteristik Model Prediksi Risiko Mortalitas Berbasis Random Forest Model Random Forest yang dikembangkan menunjukkan performa yang sangat baik dengan akurasi 87,4% dan AUC 0,91. Nilai ini berada pada kategori *excellent discrimination* sehingga menegaskan bahwa Random Forest memiliki kemampuan tinggi dalam memisahkan pasien berisiko meninggal dan hidup secara tepat. Keunggulan Random Forest terletak pada kemampuannya menangkap hubungan nonlinier dan interaksi kompleks antarvariabel klinis yang sering tidak dapat dimodelkan dengan baik oleh pendekatan statistik klasik [5].

Sejalan dengan hasil penelitian ini, Shillan et al. melaporkan bahwa Random Forest dan metode *ensemble learning* secara konsisten memberikan performa prediksi mortalitas ICU yang lebih stabil dibandingkan regresi logistik, terutama pada dataset medis yang heterogen dan tidak seimbang [8]. Rajkomar et al. juga membuktikan bahwa model *machine learning* berbasis rekam medis mampu meningkatkan ketepatan prediksi risiko kematian dan

mempercepat identifikasi pasien kritis di rumah sakit [7].

Hasil *confusion matrix* pada penelitian ini memperlihatkan tingkat *false negative* yang relatif rendah, sehingga model dinilai aman untuk implementasi klinis karena mampu meminimalkan keterlambatan deteksi pasien berisiko tinggi. Hal ini konsisten dengan temuan Nugroho et al. yang menyatakan bahwa kesalahan *false negative* merupakan aspek kritis dalam sistem prediksi mortalitas karena berimplikasi langsung terhadap keselamatan pasien [9].

## 2. Faktor Klinis yang Paling Berpengaruh terhadap Risiko Mortalitas

Hasil *feature importance* Random Forest mengidentifikasi usia lanjut, saturasi oksigen, tekanan darah sistolik, kadar hemoglobin, dan riwayat komorbid sebagai lima faktor klinis utama mortalitas pasien rawat inap. Temuan ini sejalan dengan laporan Handayani et al. yang menunjukkan bahwa usia dan parameter vital merupakan prediktor paling dominan terhadap mortalitas pasien rumah sakit di Indonesia [10].

Saturasi oksigen dan tekanan darah sistolik mencerminkan stabilitas hemodinamik pasien dan berperan penting dalam mendeteksi kondisi hipoksia dan syok, yang secara signifikan meningkatkan risiko kematian [11]. Kadar hemoglobin berhubungan langsung dengan kemampuan transportasi oksigen, sedangkan komorbiditas meningkatkan beban fisiologis pasien sehingga memperburuk prognosis klinis [12]. Dengan demikian, variabel-variabel ini valid digunakan sebagai komponen utama dalam pembentukan skor risiko mortalitas individual.

## 3. Model sebagai Dasar Sistem Peringatan Dini Klinis

Model yang dikembangkan mengelompokkan pasien ke dalam tiga kategori risiko mortalitas (<10%, 10-30%, >30%). Pendekatan klasifikasi risiko bertingkat ini sejalan dengan konsep *clinical early warning system* yang menekankan stratifikasi risiko untuk memprioritaskan intervensi medis [13].

Penelitian oleh Churpek et al. menunjukkan bahwa sistem peringatan dini berbasis *machine learning* mampu meningkatkan deteksi dini deteriorasi

klinis pasien rawat inap dan menurunkan kejadian keterlambatan penanganan [14]. Oleh karena itu, integrasi model Random Forest ini ke dalam sistem informasi rumah sakit dalam bentuk dashboard klinis real-time sangat relevan dan berpotensi besar meningkatkan keselamatan pasien dan mutu layanan.

### SIMPULAN

Model prediksi mortalitas pasien rawat inap berbasis Random Forest terbukti memiliki kinerja sangat baik dengan akurasi 87,4% dan AUC 0,91, sehingga reliabel digunakan sebagai alat prediksi dini risiko kematian. Faktor klinis utama yang paling berpengaruh terhadap mortalitas meliputi usia, saturasi oksigen, tekanan darah sistolik, kadar hemoglobin, dan riwayat komorbid. Model ini mampu mengelompokkan pasien ke dalam tingkat risiko rendah, sedang, dan tinggi, sehingga layak dijadikan dasar sistem peringatan dini klinis untuk meningkatkan ketepatan deteksi dan keselamatan pasien di rumah sakit.

### UCAPAN TERIMA KASIH

Penulis menyampaikan ucapan terima kasih kepada seluruh pihak yang telah memberikan dukungan dan kontribusi dalam penyusunan penelitian ini. Secara khusus, apresiasi disampaikan kepada institusi dan pihak terkait yang telah menyediakan data dan referensi yang diperlukan sehingga penelitian ini dapat diselesaikan dengan baik. Penulis juga mengucapkan terima kasih kepada para reviewer dan editor atas masukan, saran, dan kritik konstruktif yang sangat bermanfaat dalam penyempurnaan artikel ini. Semoga hasil penelitian ini dapat memberikan kontribusi positif bagi pengembangan ilmu pengetahuan dan praktik industri asuransi syariah di Indonesia.

### DAFTAR PUSTAKA

- [1] Kementerian Kesehatan Republik Indonesia, Profil Kesehatan Indonesia 2023. Jakarta, Indonesia: Kemenkes RI, 2024.
- [2] D. W. Hosmer, S. Lemeshow, and R. X. Sturdivant, Applied Logistic Regression, 3rd ed. New York, NY, USA: John Wiley & Sons, 2013.
- [3] F. E. Harrell, Regression Modeling Strategies, 2nd ed. New York, NY, USA: Springer, 2015.
- [4] Z. Obermeyer and E. J. Emanuel, "Predicting the future—Big data,

- machine learning, and clinical medicine,” *New England Journal of Medicine*, vol. 375, no. 13, pp. 1216–1219, 2016, doi: 10.1056/NEJMp1606181.
- [5] L. Breiman, “Random forests,” *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.
- [6] A. E. W. Johnson et al., “MIMIC-III, a freely accessible critical care database,” *Scientific Data*, vol. 3, p. 160035, 2016, doi: 10.1038/sdata.2016.35.
- [7] A. Rajkomar et al., “Scalable and accurate deep learning with electronic health records,” *NPJ Digital Medicine*, vol. 1, p. 18, 2018, doi: 10.1038/s41746-018-0029-1.
- [8] D. Shillan, J. A. C. Sterne, A. Champneys, and B. Gibbison, “Use of machine learning to analyse routinely collected intensive care unit data: A systematic review,” *Critical Care*, vol. 23, p. 284, 2019.
- [9] A. Nugroho, Y. Kurniawan, and E. Setiawan, “Prediction of inpatient mortality using Cox proportional hazard model in Indonesian hospital,” *Journal of Medical Systems*, vol. 45, p. 12, 2021.
- [10] P. W. Handayani et al., “Hospital mortality prediction using logistic regression in Indonesian hospitals,” *BMC Medical Informatics and Decision Making*, vol. 20, p. 302, 2020.
- [11] J.-L. Vincent and D. De Backer, “Circulatory shock,” *New England Journal of Medicine*, vol. 369, no. 18, pp. 1726–1734, 2013.
- [12] G. Weiss and L. T. Goodnough, “Anemia of chronic disease,” *New England Journal of Medicine*, vol. 352, no. 10, pp. 1011–1023, 2005.
- [13] C. P. Subbe et al., “Validation of a modified Early Warning Score,” *QJM: An International Journal of Medicine*, vol. 94, no. 10, pp. 521–526, 2001.
- [14] M. M. Churpek, T. C. Yuen, and D. P. Edelson, “Risk stratification of hospitalized patients using electronic health record data,” *Chest*, vol. 143, no. 6, pp. 175–182, 2013.