



Model Klasifikasi Usia Kematian Neonatal (≤ 1 Hari dan > 1 Hari) Berbasis *Machine Learning*

Nugraheni Kusumawati¹, Hugi Cerlyawati², Yanita Sri Mulyani³, Evina Widianawati^{4,*}

¹⁻⁴Rekam Medis dan Informasi Kesehatan, Fakultas Kesehatan, Universitas Dian Nuswantoro, Semarang, Indonesia

Informasi Artikel

Sejarah Artikel:
 Submit: 18 Februari 2026
 Revisi: 01 Maret 2026
 Diterima: 04 Maret 2026
 Diterbitkan: 24 Maret 2026

Kata Kunci

Prediksi, Neonatal, Usia, Model, Machine Learning

Korespondensi

E-mail: evina.widianawati@dsn.dinus.ac.id*

A B S T R A C T

The neonatal period, particularly within the first 24 hours of life, represents a critical phase with high risks of morbidity and mortality. Identifying clinical patterns that distinguish neonates aged ≤ 1 day from those aged > 1 day is essential to support data-driven clinical decision-making. This study aimed to compare the performance of several machine learning algorithms in classifying neonatal age and to identify the most contributory clinical factors. This observational analytic study utilized secondary medical record data of neonates from Hospital X in Semarang City. Predictor variables included sex, gestational age, birth weight, body length, number of secondary diagnoses, primary diagnosis, and mode of delivery. Five algorithms were evaluated: Logistic Regression, Support Vector Machine (RBF), Random Forest, HistGradientBoosting, and XGBoost. Model performance was assessed using Stratified K-Fold Cross-Validation with accuracy, macro precision, macro recall, and F1 score as evaluation metrics. Feature importance analysis was conducted using the Random Forest model. Random Forest demonstrated the best performance, achieving an accuracy of 0.686 and a weighted F1 score of 0.610 ± 0.182 . HistGradientBoosting and SVM both achieved an accuracy of 0.657, while Logistic Regression and XGBoost achieved accuracies of 0.595 and 0.557, respectively. Gestational age was identified as the most influential factor, followed by body length, number of secondary diagnoses, and birth weight. Several primary diagnoses related to respiratory disorders were also among the important predictors. Random Forest provided the best balance of accuracy and stability in predicting neonatal age at death.

Abstrak

Periode neonatal, khususnya dalam 24 jam pertama kehidupan, merupakan fase kritis dengan risiko morbiditas dan mortalitas tinggi. Identifikasi pola klinis yang membedakan neonatus usia ≤ 1 hari dan > 1 hari penting untuk mendukung pengambilan keputusan klinis berbasis data. Tujuan penelitian ini yaitu membandingkan performa beberapa algoritma *machine learning* dalam mengklasifikasikan usia neonatus serta mengidentifikasi faktor klinis yang paling berkontribusi. Penelitian analitik observasional ini menggunakan data sekunder rekam medis neonatus sebanyak 41 kasus di rumah sakit X di kota Semarang. Variabel prediktor meliputi jenis kelamin, usia gestasi, berat badan, tinggi badan, jumlah diagnosa sekunder, diagnosa utama, dan cara melahirkan. Lima algoritma diuji: *Logistic Regression*, *Support Vector Machine (RBF)*, *Random Forest*, *HistGradientBoosting*, dan *XGBoost*. Evaluasi dilakukan menggunakan *Stratified K-Fold Cross Validation* dengan metrik akurasi, presisi makro, *recall* makro, dan *F1 score*. Analisis faktor terpenting dilakukan menggunakan *Random Forest*. *Random Forest* menunjukkan performa terbaik dengan akurasi 0,686 dan *F1* tertimbang $0,610 \pm 0,182$. *HistGradientBoosting* dan *SVM* mencapai akurasi 0,657, sedangkan

Regresi Logistik dan *XGBoost* masing-masing 0,595 dan 0,557. Usia gestasi merupakan faktor paling berpengaruh, diikuti tinggi badan, jumlah diagnosa sekunder, dan berat badan. Beberapa diagnosa utama terkait gangguan pernapasan juga termasuk dalam faktor penting. *Random Forest* memberikan kombinasi akurasi dan stabilitas terbaik dalam memprediksi usia kematian bayi baru lahir.

This is an open access article under the CC-BY-SA license



1. Pendahuluan

Periode *neonatal*, khususnya dalam 24 jam pertama kehidupan, merupakan fase kritis yang menentukan stabilitas fisiologis bayi setelah lahir. Adaptasi sistem respirasi, kardiovaskular, dan metabolik terjadi secara cepat pada periode ini, dan kegagalan adaptasi dapat meningkatkan risiko morbiditas maupun mortalitas [1,2]. Secara global, proporsi kematian neonatal tertinggi terjadi pada hari pertama kehidupan, yang mencerminkan tingginya kerentanan pada fase awal tersebut [3]. Prematuritas dan berat badan lahir rendah telah lama diidentifikasi sebagai determinan utama luaran klinis neonatal [4,5]. Selain itu, gangguan pernapasan serta komplikasi obstetrik juga berkontribusi signifikan terhadap ketidakstabilan kondisi bayi baru lahir [6]. Oleh karena itu, identifikasi pola klinis yang dapat membedakan kondisi neonatus pada usia ≤ 1 hari dan >1 hari menjadi penting dalam mendukung pengambilan keputusan klinis dan strategi pemantauan dini.

Dalam beberapa tahun terakhir, pendekatan *machine learning* semakin luas digunakan dalam bidang kesehatan untuk membantu klasifikasi dan prediksi berbasis data klinis [7,8]. Model seperti *Logistic Regression*, *Support Vector Machine (SVM)*, *Random Forest*, dan *XGBoost* memungkinkan analisis hubungan non-linear serta interaksi kompleks antar variabel yang sulit ditangkap oleh metode statistik konvensional [9–11]. Pendekatan ensemble berbasis pohon keputusan, khususnya *Random Forest* dan *gradient boosting*, telah menunjukkan performa prediksi yang kompetitif dalam berbagai studi klinis [12,13]. Keunggulan pendekatan ini terletak pada kemampuannya mengintegrasikan variabel numerik dan kategorik secara simultan serta menghasilkan model yang dapat dievaluasi melalui berbagai metrik performa. Namun demikian, pada dataset dengan ukuran sampel terbatas, evaluasi performa model perlu dilakukan secara hati-hati karena variasi antar *fold* validasi dapat meningkat dan memengaruhi stabilitas estimasi [14].

Meskipun berbagai faktor risiko neonatal telah banyak diteliti secara epidemiologis, integrasi berbagai variabel klinis dalam satu kerangka analitik berbasis komputasi masih relatif terbatas, khususnya dalam konteks rumah sakit di tingkat lokal. Pendekatan berbasis data memungkinkan identifikasi pola tersembunyi yang tidak selalu terlihat melalui analisis univariat atau bivariat konvensional. Dengan memanfaatkan data rekam medis sebagai sumber informasi *real-world evidence*, analisis *machine learning* berpotensi memberikan wawasan yang lebih komprehensif mengenai interaksi antara faktor maturitas kehamilan, kondisi klinis, dan faktor obstetrik dalam periode neonatal dini. Oleh karena itu, pengembangan model klasifikasi yang tidak hanya akurat tetapi juga interpretatif menjadi relevan untuk mendukung praktik klinis berbasis bukti dan perencanaan intervensi preventif pada kelompok risiko tinggi. Meskipun *machine learning* telah banyak digunakan dalam penelitian neonatal, studi yang membandingkan beberapa algoritma menggunakan data klinis rutin di rumah sakit tingkat lokal di Indonesia masih terbatas.

Sejalan dengan latar belakang tersebut, penelitian ini mengintegrasikan pendekatan klinis dan komputasional untuk mengevaluasi kemampuan beberapa algoritma *machine learning* dalam mengklasifikasikan usia kematian neonatal berdasarkan karakteristik biologis, obstetrik, dan diagnostik. Dengan memanfaatkan data rekam medis sebagai sumber informasi *real-world evidence*, studi ini tidak hanya menilai performa prediktif model, tetapi juga menyoroti variabel-variabel yang memiliki kontribusi terbesar dalam membedakan kelompok usia ≤ 1 hari dan >1 hari. Oleh karena itu,

penelitian ini bertujuan untuk membandingkan performa beberapa algoritma klasifikasi dalam membedakan kelompok usia neonatus (>1 hari dan ≤ 1 hari), serta mengidentifikasi faktor klinis yang paling berkontribusi dalam proses klasifikasi. Pendekatan ini diharapkan tidak hanya memberikan gambaran performa komputasional model, tetapi juga memperkuat pemahaman mengenai relevansi klinis variabel yang berperan dalam periode neonatal awal.

2. Metode Penelitian

Penelitian ini menggunakan desain analitik observasional berbasis data sekunder dari rekam medis neonatus di salah satu RS di Kota Semarang dari tahun 2017-2021. Variabel target adalah usia neonatus yang dikategorikan menjadi dua kelompok, yaitu ≤ 1 hari dan >1 hari. Variabel prediktor yang dianalisis meliputi jenis kelamin, usia gestasi (minggu), berat badan (gram), tinggi badan (cm), jumlah diagnosa sekunder, diagnosa utama, serta cara melahirkan. Kolom jumlah tindakan tidak dimasukkan dalam analisis untuk menghindari potensi bias prosedural. Tahap praproses data dilakukan melalui pembersihan dan transformasi variabel. Variabel kategorik dikonversi menggunakan teknik one-hot encoding, sedangkan variabel target dikodekan secara numerik melalui label *encoding*, sesuai praktik umum dalam pengembangan model prediksi klinis [15].

Penelitian ini membandingkan lima algoritma klasifikasi, yaitu *Logistic Regression*, *Support Vector Machine* (kernel RBF), *Random Forest*, *HistGradientBoosting*, dan *XGBoost*. *Logistic Regression* merepresentasikan pendekatan linear, SVM mewakili pendekatan berbasis kernel, sedangkan *Random Forest*, *HistGradientBoosting*, dan *XGBoost* termasuk dalam pendekatan *ensemble learning* berbasis pohon keputusan [12,16,17]. Evaluasi performa model dilakukan menggunakan *Stratified K-Fold Cross Validation* dengan jumlah *fold* yang disesuaikan berdasarkan distribusi minimum tiap kelas (2-5 fold) untuk mempertahankan proporsi distribusi kelas pada setiap *fold* validasi [14,18]. Seluruh model dijalankan menggunakan parameter yang ditetapkan dalam skrip analisis tanpa prosedur *hyperparameter tuning* tambahan. Metrik evaluasi meliputi akurasi, presisi makro, *recall* makro, F1 makro, dan F1 tertimbang, yang direkomendasikan dalam evaluasi model klasifikasi medis untuk menilai keseimbangan performa antar kelas [19]. Selain nilai rata-rata, standar deviasi dihitung untuk mengukur stabilitas model. Analisis faktor terpenting dilakukan menggunakan pendekatan *feature importance* pada *Random Forest* untuk mengidentifikasi kontribusi relatif tiap variabel dalam proses klasifikasi [20].

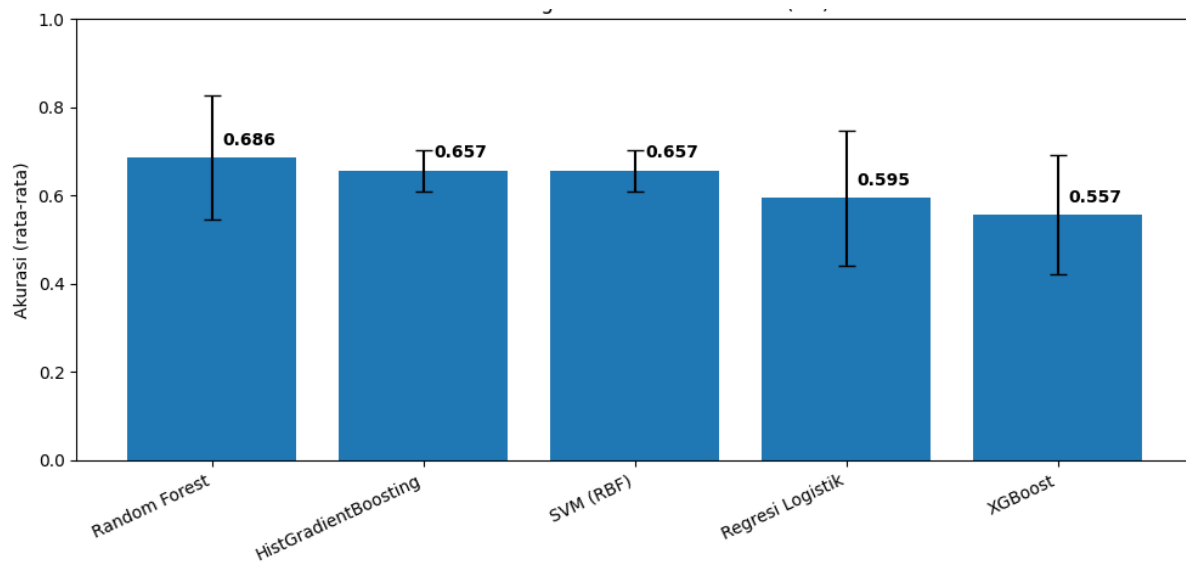
3. Hasil dan Pembahasan

Tabel 1. Hasil analisis deskriptif karakteristik neonatal

Variabel		Usia Kematian f(%)		Total
		> 1 hari	≤ 1 hari	
Jenis Kelamin	Laki-laki	8 (20%)	15 (37%)	23 (56%)
	Perempuan	8 (20%)	10 (24%)	18 (44%)
Cara Melahirkan	Sectio Caesarea	8 (20%)	10 (24%)	18 (44%)
	Spontan	7 (17%)	15 (37%)	22 (54%)
	NA	1 (2%)		1 (2%)
Diganosa Utama	Severe birth asphyxia	1 (2%)	5 (12%)	6 (15%)
	Respiratory failure of newborn	1 (2%)	4 (10%)	5 (12%)
	Respiratory distress of newborn, unspecified	1 (2%)	3 (7%)	4 (10%)
	Other preterm infants	2(5%)	2(5%)	4 (10%)
	Multiple congenital malformations, not elsewhere classified		3 (7%)	3 (7%)
	Extreme immaturity	1 (2%)	1 (2%)	2(5%)
	Respiratory distress syndrome of newborn	2(5%)		2(5%)
	Other low birth weight	1 (2%)		1 (2%)
	Neonatal aspiration of meconium		1 (2%)	1 (2%)

	Persistent fetal circulation	1 (2%)	1 (2%)
Total	16 (39%)	25 (61%)	41 (100%)

Berdasarkan Tabel 1, dari total 41 kasus kematian neonatal, sebagian besar terjadi pada usia ≤ 1 hari yaitu 25 kasus (61%), sedangkan usia >1 hari sebanyak 16 kasus (39%). Distribusi jenis kelamin menunjukkan bahwa kematian pada bayi laki-laki sedikit lebih tinggi (23 kasus; 56%) dibandingkan perempuan (18 kasus; 44%). Berdasarkan cara melahirkan, persalinan spontan merupakan proporsi terbesar (22 kasus; 54%), diikuti *sectio caesarea* (18 kasus; 44%), dan sebagian kecil tidak tercatat (2%). Dari sisi diagnosa utama, gangguan pernapasan mendominasi penyebab kematian, terutama *severe birth asphyxia* (15%), *respiratory failure of newborn* (12%), dan *respiratory distress of newborn, unspecified* (10%). Selain itu, kondisi terkait prematuritas seperti *other preterm infants* (10%), *extreme immaturity* (5%), serta kelainan kongenital juga berkontribusi terhadap kematian neonatal. Secara keseluruhan, pola ini menunjukkan bahwa mayoritas kematian terjadi pada periode sangat dini (≤ 1 hari) dan banyak berkaitan dengan gangguan respirasi serta komplikasi prematuritas.



Gambar 1. Perbandingan akurasi prediksi usia neonatal

Hasil pengujian menunjukkan bahwa kelima algoritma memiliki kemampuan yang berbeda dalam membedakan kelompok usia neonatus (>1 hari dan ≤ 1 hari). Model dengan hasil terbaik adalah *Random Forest* dengan akurasi 0,686. Artinya, sekitar 68,6% data dapat diklasifikasikan dengan benar oleh model ini. Di bawahnya terdapat *HistGradientBoosting* dan *SVM (RBF)* yang masing-masing memiliki akurasi 0,657. Regresi Logistik menghasilkan akurasi 0,595, sedangkan *XGBoost* memiliki akurasi paling rendah yaitu 0,557. Dari hasil ini dapat disimpulkan bahwa *Random Forest* lebih mampu mengenali pola dalam data dibandingkan model lainnya pada dataset yang digunakan.

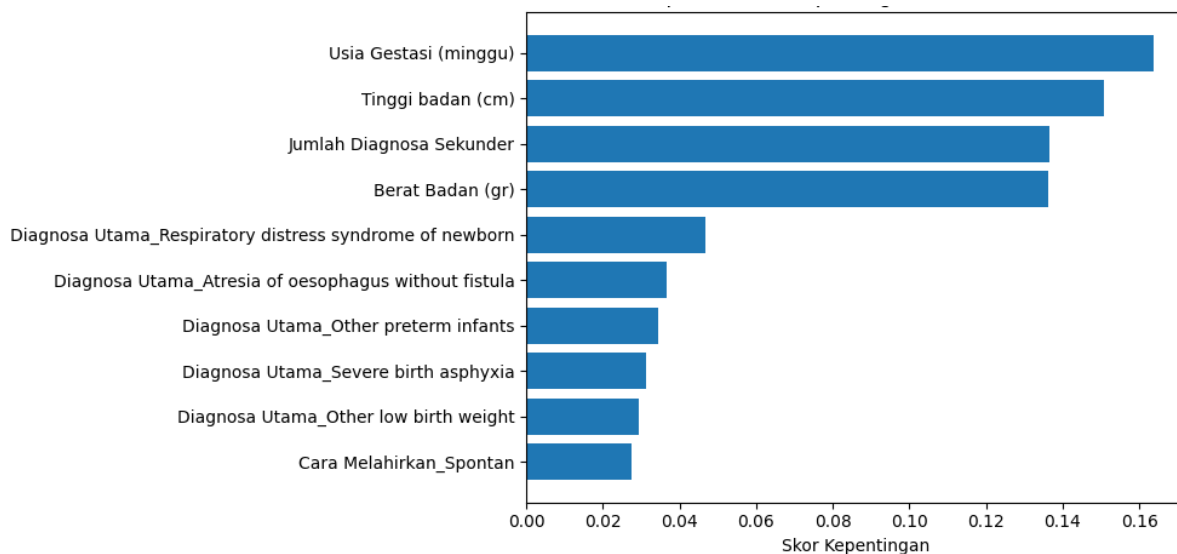
Jika dilihat dari grafik akurasi beserta error bar (standar deviasi), *Random Forest* juga menunjukkan performa yang relatif stabil dibandingkan beberapa model lainnya. Meskipun terdapat variasi antar *fold* validasi, nilai rata-rata akurasi tetap yang tertinggi. *HistGradientBoosting* dan *SVM* memiliki akurasi yang hampir sama serta rentang variasi yang lebih kecil, namun keduanya masih berada di bawah *Random Forest*. Regresi Logistik dan *XGBoost* tidak hanya memiliki akurasi lebih rendah, tetapi juga menunjukkan variasi performa yang cukup besar pada beberapa *fold*, yang menandakan bahwa model tersebut kurang konsisten pada dataset ini.

Tabel 2. Perbandingan evaluasi performa *machine learning*

Algoritma	Presisi Makro (rata-rata±SD)	Recall Makro (rata-rata±SD)	F1 Makro (rata-rata±SD)	F1 Tertimbang (rata-rata±SD)
<i>Random Forest</i>	0.547 ± 0.295	0.575 ± 0.150	0.533 ± 0.205	0.610 ± 0.182
<i>HistGradientBoosting</i>	0.329 ± 0.023	0.500 ± 0.000	0.396 ± 0.017	0.522 ± 0.058
SVM (RBF)	0.329 ± 0.023	0.500 ± 0.000	0.396 ± 0.017	0.522 ± 0.058
Regresi Logistik	0.480 ± 0.254	0.505 ± 0.162	0.479 ± 0.196	0.552 ± 0.180
XGBoost	0.495 ± 0.191	0.518 ± 0.192	0.490 ± 0.173	0.545 ± 0.143

Jika dilihat dari ukuran evaluasi lain seperti presisi, *recall*, dan *F1 score*, hasilnya juga menunjukkan pola yang serupa. *Random Forest* memiliki nilai F1 tertimbang sebesar 0,610, yang merupakan nilai tertinggi dibandingkan algoritma lain. Ini menunjukkan bahwa model tersebut tidak hanya akurat, tetapi juga cukup seimbang dalam memprediksi kedua kelompok usia. Sementara itu, *HistGradientBoosting* dan SVM memiliki nilai presisi yang lebih rendah, meskipun kemampuan mendeteksi kelas (*recall*) berada pada tingkat sedang. Regresi Logistik dan XGBoost menunjukkan performa yang cukup stabil, tetapi tetap belum melebihi *Random Forest*. Dengan kata lain, secara keseluruhan *Random Forest* memberikan kombinasi performa yang paling baik di antara semua model yang diuji. Nilai standar deviasi pada setiap metrik menunjukkan adanya variasi performa antar pembagian data (fold) selama proses validasi silang. *Random Forest* memiliki variasi yang masih dalam batas wajar, sedangkan beberapa model lain menunjukkan fluktuasi yang lebih besar. Hal ini kemungkinan dipengaruhi oleh jumlah data yang relatif terbatas, sehingga perubahan kecil dalam pembagian data dapat memengaruhi hasil evaluasi. Oleh karena itu, meskipun *Random Forest* menunjukkan performa terbaik, hasil ini tetap perlu ditafsirkan dengan hati-hati.

Berdasarkan tabel evaluasi presisi, *recall*, dan *F1 score*, *Random Forest* juga memiliki nilai F1 tertimbang tertinggi (0,610 ± 0,182), yang menunjukkan keseimbangan yang lebih baik dalam memprediksi kedua kelas usia. Sementara itu, *HistGradientBoosting* dan SVM memiliki presisi makro yang lebih rendah, meskipun nilai *recall* berada pada tingkat sedang. Hal ini menunjukkan bahwa kedua model tersebut cukup mampu mendeteksi kelas, tetapi masih kurang tepat dalam beberapa prediksi. Secara keseluruhan, kombinasi akurasi tertinggi, *F1 score* yang lebih baik, serta stabilitas performa membuat *Random Forest* menjadi model yang paling optimal dalam penelitian ini.



Gambar 2. Top 10 faktor terpenting menggunakan *Random Forest*

Analisis faktor terpenting dari model *Random Forest* menunjukkan bahwa Usia Gestasi (minggu) merupakan faktor yang paling berpengaruh dalam menentukan kelompok usia neonatus. Nilai skor kepentingannya paling tinggi dibandingkan variabel lain, yang menunjukkan bahwa kematangan kehamilan memiliki peran utama dalam membedakan bayi dengan usia ≤ 1 hari dan > 1 hari. Setelah itu, faktor yang juga memiliki pengaruh besar adalah Tinggi Badan (cm), Jumlah Diagnosa Sekunder, dan Berat Badan (gram). Ketiga variabel ini termasuk dalam kelompok faktor dengan skor kepentingan tinggi dan menunjukkan bahwa parameter fisik serta kompleksitas kondisi klinis bayi sangat berperan dalam proses klasifikasi.

Beberapa kategori dalam Diagnosa Utama juga muncul dalam sepuluh besar faktor terpenting, antara lain *respiratory distress syndrome of newborn*, *atresia of oesophagus without fistula*, *other preterm infants*, *severe birth asphyxia*, dan *other low birth weight*. Munculnya beberapa diagnosa ini menunjukkan bahwa kondisi pernapasan dan komplikasi prematuritas menjadi penanda penting dalam membedakan kelompok usia. Selain itu, variabel Cara Melahirkan_Spontaneus juga termasuk dalam sepuluh faktor teratas, meskipun skor kepentingannya lebih rendah dibandingkan faktor fisiologis seperti usia gestasi dan ukuran tubuh bayi. Secara keseluruhan, pola ini menunjukkan bahwa model *Random Forest* tidak hanya mengandalkan satu jenis variabel, tetapi memanfaatkan kombinasi faktor biologis (usia gestasi dan ukuran tubuh), faktor klinis (diagnosa utama dan sekunder), serta faktor obstetrik (cara melahirkan). Hal ini memperkuat bahwa klasifikasi usia neonatus dalam penelitian ini dipengaruhi oleh interaksi antara tingkat kematangan kehamilan dan kondisi klinis bayi saat lahir.

Berdasarkan hasil evaluasi, *Random Forest* menunjukkan kombinasi performa yang paling baik dibandingkan algoritma lainnya, tidak hanya dari sisi akurasi tetapi juga dari keseimbangan presisi, *recall*, dan *F1 score*. Nilai *F1* tertimbang sebesar $0,610 \pm 0,182$ menunjukkan bahwa model ini relatif lebih stabil dalam memprediksi kedua kelompok usia secara seimbang. Keunggulan ini dapat dijelaskan oleh karakteristik ensemble learning yang mengurangi variansi melalui agregasi banyak pohon keputusan dan meningkatkan stabilitas prediksi [12,21]. Studi metodologis sebelumnya menunjukkan bahwa model berbasis pohon cenderung lebih adaptif terhadap data klinis heterogen dibandingkan model linear tradisional [22] serta metode ini efektif dalam menangani data dengan hubungan non-linear dan interaksi kompleks antar variabel. Sebaliknya, Regresi Logistik memiliki keterbatasan dalam menangkap hubungan *non-linear* antar variabel, sedangkan SVM memerlukan pemilihan kernel dan parameter yang tepat untuk mencapai performa optimal [9,23]. *XGBoost* sebagai metode boosting berbasis gradient seringkali menunjukkan performa optimal pada dataset besar, namun pada dataset kecil dapat mengalami fluktuasi performa yang lebih tinggi akibat sensitivitas terhadap pembagian data [17]. Hal ini konsisten dengan literatur yang menyatakan bahwa estimasi performa pada sampel kecil dapat menghasilkan variasi antar *fold* yang signifikan [14,18].

Perbedaan nilai presisi dan *recall* antar model juga memberikan gambaran karakteristik masing-masing algoritma. *HistGradientBoosting* dan SVM menunjukkan *recall* yang moderat (0,500), namun presisi yang relatif rendah, yang mengindikasikan adanya prediksi positif yang kurang tepat. *Random Forest* menunjukkan keseimbangan yang lebih baik antara kemampuan mendeteksi kelas dan ketepatan prediksi. Namun demikian, standar deviasi yang masih terlihat pada beberapa metrik menunjukkan bahwa ukuran sampel yang relatif kecil dapat memengaruhi stabilitas model. Studi metodologis sebelumnya juga menekankan bahwa pada dataset kecil, variasi antar *fold* validasi silang dapat meningkat dan berpotensi memengaruhi estimasi performa model [14].

Dari perspektif klinis, faktor terpenting yang muncul dalam model *Random Forest* sangat konsisten dengan literatur neonatologi. Pendekatan interpretabilitas berbasis feature importance memungkinkan identifikasi variabel yang memiliki kontribusi terbesar terhadap hasil prediksi [20,24]. Dalam konteks pengembangan model prediksi klinis, transparansi dan interpretabilitas merupakan aspek penting

untuk memastikan penerimaan dan kepercayaan dalam praktik medis [19,25]. Usia gestasi sebagai faktor dengan kontribusi tertinggi mencerminkan tingkat maturitas organ bayi, terutama sistem respirasi dan neurologis, yang sangat menentukan stabilitas kondisi neonatal awal. Prematuritas telah lama diketahui sebagai determinan utama morbiditas dan mortalitas neonatal [5]. Berat badan lahir dan parameter antropometri lainnya juga merupakan indikator penting status pertumbuhan intrauterin dan risiko komplikasi. Munculnya diagnosa seperti *respiratory distress syndrome* dan *severe birth asphyxia* sebagai faktor penting semakin memperkuat bahwa gangguan pernapasan dan komplikasi hipoksia memiliki peran besar pada periode neonatal dini. Sebagai pengembangan lanjutan, penggunaan metode interpretabilitas yang lebih komprehensif seperti *SHAP (Shapley Additive Explanations)* direkomendasikan untuk meningkatkan transparansi model dan memberikan pemahaman yang lebih mendalam mengenai kontribusi masing-masing variabel terhadap prediksi.

Secara keseluruhan, temuan ini menunjukkan bahwa model *machine learning* tidak hanya bekerja secara statistik, tetapi juga mampu merefleksikan pola klinis yang telah diidentifikasi dalam penelitian epidemiologis sebelumnya. Kombinasi faktor maturitas kehamilan, kondisi fisik bayi, serta kompleksitas diagnosa klinis berkontribusi dalam membedakan kelompok usia neonatus. Namun demikian, mengingat keterbatasan ukuran sampel dan variasi performa antar *fold*, diperlukan penelitian lanjutan dengan jumlah data yang lebih besar serta validasi eksternal untuk memastikan bahwa model ini dapat diterapkan secara lebih luas dalam konteks klinis.

4. Kesimpulan

Penelitian ini menunjukkan bahwa pendekatan *machine learning* dapat digunakan untuk membedakan kelompok usia neonatus (≤ 1 hari dan > 1 hari) berdasarkan karakteristik klinis dan obstetrik. Dari lima algoritma yang diuji, *Random Forest* memberikan performa terbaik dengan kombinasi akurasi, *F1 score*, dan stabilitas yang lebih unggul dibandingkan model lainnya. Faktor yang paling berpengaruh dalam klasifikasi adalah usia gestasi, diikuti oleh tinggi badan, jumlah diagnosa sekunder, berat badan, serta beberapa diagnosa utama terkait gangguan pernapasan dan prematuritas. Temuan ini konsisten dengan literatur klinis yang menegaskan pentingnya maturitas kehamilan dan kondisi respirasi dalam periode neonatal awal. Meskipun demikian, variasi performa antar *fold* menunjukkan bahwa ukuran sampel yang terbatas dapat memengaruhi stabilitas model, sehingga diperlukan penelitian lanjutan dengan data yang lebih besar dan validasi eksternal untuk meningkatkan generalisasi dan potensi implementasi klinis. Model yang dikembangkan berpotensi diimplementasikan sebagai alat pendukung keputusan klinis maupun sistem skrining risiko dini pada unit perawatan neonatal untuk membantu identifikasi cepat neonatus berisiko tinggi dan mendukung intervensi yang lebih tepat waktu.

Daftar Pustaka

- [1] L. Hug, M. Alexander, D. You, and L. Alkema, "National, regional, and global levels and trends in neonatal mortality between 1990 and 2017, with scenario-based projections to 2030: A systematic analysis," *Lancet Glob. Health*, vol. 7, no. 6, pp. e710–e720, 2019.
- [2] M. J. Sankar, C. K. Natarajan, R. R. Das, R. Agarwal, A. Chandrasekaran, and V. K. Paul, "When do newborns die? A systematic review of timing of overall and cause-specific neonatal deaths in developing countries," *Lancet Glob. Health*, vol. 4, no. 10, pp. e700–e711, 2016.
- [3] World Health Organization, "Newborn mortality fact sheet," Geneva, Switzerland: WHO, 2023.
- [4] R. L. Goldenberg, J. F. Culhane, J. D. Iams, and R. Romero, "Epidemiology and causes of preterm birth," *Lancet*, vol. 371, no. 9606, pp. 75–84, 2008.
- [5] H. Blencowe et al., "National, regional, and worldwide estimates of preterm birth rates in the year 2010," *Lancet*, vol. 379, no. 9832, pp. 2162–2172, 2012.
- [6] B. J. Stoll et al., "Trends in care practices, morbidity, and mortality of extremely preterm neonates, 1993–2012," *JAMA*, vol. 314, no. 10, pp. 1039–1051, 2015.
- [7] A. Rajkomar, J. Dean, and I. Kohane, "*Machine learning* in medicine," *N. Engl. J. Med.*, vol. 380, no. 14, pp. 1347–1358, 2019.

- [8] A. L. Beam and I. S. Kohane, "Big data and *machine learning* in health care," *JAMA*, vol. 319, no. 13, pp. 1317–1318, 2018.
- [9] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.
- [10] A. Esteva et al., "A guide to deep learning in healthcare," *Nat. Med.*, vol. 25, no. 1, pp. 24–29, 2019.
- [11] E. J. Topol, "High-performance medicine: The convergence of human and artificial intelligence," *Nat. Med.*, vol. 25, no. 1, pp. 44–56, 2019.
- [12] L. Breiman, "*Random Forests*," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
- [13] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *Ann. Stat.*, vol. 29, no. 5, pp. 1189–1232, 2001.
- [14] G. Varoquaux, "Cross-validation failure: Small sample sizes lead to large error bars," *NeuroImage*, vol. 180, pp. 68–77, 2018.
- [15] E. W. Steyerberg, *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating*, 2nd ed. Cham, Switzerland: Springer, 2019.
- [16] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, 2016, pp. 785–794.
- [17] G. Biau and E. Scornet, "A *Random Forest* guided tour," *Test*, vol. 25, no. 2, pp. 197–227, 2016.
- [18] J. Lever, M. Krzywinski, and N. Altman, "Points of significance: Cross-validation," *Nat. Methods*, vol. 13, no. 8, pp. 703–704, 2016.
- [19] G. S. Collins, J. B. Reitsma, D. G. Altman, and K. G. M. Moons, "Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD statement," *BMJ*, vol. 350, p. g7594, 2015.
- [20] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Adv. Neural Inf. Process. Syst. (NeurIPS)*, vol. 30, 2017, pp. 4765–4774.
- [21] T. G. Dietterich, "Ensemble methods in machine learning," in *Multiple Classifier Systems, Lecture Notes in Computer Science*, vol. 1857, 2000, pp. 1–15.
- [22] R. Caruana and A. Niculescu-Mizil, "An empirical comparison of supervised learning algorithms," in *Proc. 23rd Int. Conf. Mach. Learn. (ICML)*, 2006, pp. 161–168.
- [23] W. S. Noble, "What is a support vector machine?" *Nat. Biotechnol.*, vol. 24, no. 12, pp. 1565–1567, 2006.
- [24] C. Molnar, *Interpretable Machine Learning*, 2nd ed., 2022. [Online]. Available: <https://christophm.github.io/interpretable-ml-book/>
- [25] L. Wynants et al., "Prediction models for diagnosis and prognosis of COVID-19 infection: Systematic review and critical appraisal," *BMJ*, vol. 369, p. m1328, 2020.