RESEARCH ARTICLE • OPEN ACCESS

Evaluasi Kinerja Model *Random Forest* dan *LightGBM* untuk Klasifikasi Status Imunisasi Hepatitis B (HB-0) pada Balita

Ummul Auliyah Syam, dkk.

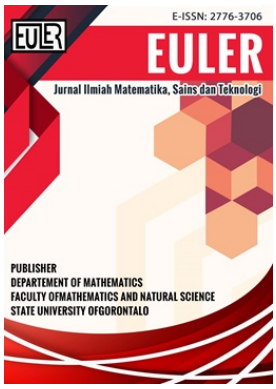
Volume 13, Issue 1, Pages 1–8, April 2025

Diterima 9 Januari 2025, Direvisi 14 Maret 2025, Disetujui 26 Maret 2025, Diterbitkan 1 April 2025

To Cite this Article : U. A. Syam, I. Irdayanti, I. Magfirrah, B. Sartono, dan A. R. Firdawanti,"Evaluasi Kinerja Model *Random Forest* dan *LightGBM* untuk Klasifikasi Status Imunisasi Hepatitis B (HB-0) pada Balita", *Euler J. Ilm. Mat. Sains dan Teknol.*, vol. 13, no. 1, pp. 1–8, 2025, *https://doi.org/10.37905/euler.v13i1.29762*

© 2025 by author(s)

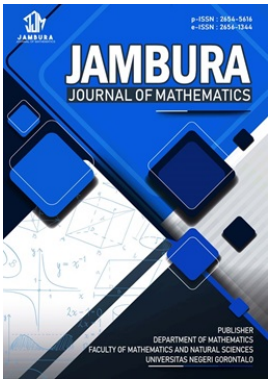
JOURNAL INFO • EULER : JURNAL ILMIAH MATEMATIKA, SAINS DAN TEKNOLOGI



Homepage	:	http://ejurnal.ung.ac.id/index.php/euler/index
Journal Abbreviation	:	Euler J. Ilm. Mat. Sains dan Teknol.
Frequency	:	Three times a year
Publication Language	:	English (preferable), Indonesia
DOI	:	https://doi.org/10.37905/euler
Online ISSN	:	2776-3706
License	:	Creative Commons Attribution-NonCommercial 4.0 International License
Publisher	:	Department of Mathematics, Universitas Negeri Gorontalo
Country	:	Indonesia
OAI Address	:	http://ejurnal.ung.ac.id/index.php/euler/oai
Google Scholar ID	:	QF_r-_gAAAAJ
Email	:	euler@ung.ac.id

JAMBURA JOURNAL • FIND OUR OTHER JOURNALS

Jambura Journal of Biomathematics



Jambura Journal of Mathematics

Jambura Journal of Mathematics Education

Jambura Journal of Probability and Statistics

Euler : Jurnal Ilmiah Matematika, Sains dan Teknologi, Volume 13, Issue 1, Pages 1–8, April 2025 https://doi.org/10.37905/euler.v13i1.29762

Research Article

Check for updates

Evaluasi Kinerja Model *Random Forest* dan *LightGBM* untuk Klasifikasi Status Imunisasi Hepatitis B (HB-0) pada Balita

Ummul Auliyah Syam[1,*], Irdayanti[1], Indah Magfirrah[1], Bagus Sartono[1], Aulia Rizki Firdawanti[1]

[1]*Program Studi Statistika dan Sains Data, IPB University, Bogor, Indonesia*

ARTICLE HISTORY
Diterima 9 Januari 2025
Direvisi 14 Maret 2025
Disetujui 26 Maret 2025
Diterbitkan 1 April 2025

KATA KUNCI
Interpretabilitas
Klasifikasi
SHAP

KEYWORDS
Classification
Interpretability
SHAP

ABSTRAK. *Imunisasi hepatitis B (HB-0) pada balita adalah langkah penting untuk mencegah penularan hepatitis B sejak dini dan meningkatkan kesehatan masyarakat. Penelitian ini bertujuan untuk mengklasifikasikan status imunisasi HB-0 pada balita di Provinsi Jawa Barat. Metode yang digunakan adalah algoritma Random Forest dan LightGBM. Hasil penelitian menunjukkan bahwa model Random Forest memiliki balanced accuracy sebesar 0,8443 yang sedikit lebih tinggi dari LightGBM (0,8357). Hal ini menunjukkan bahwa Random Forest memiliki kinerja yang lebih baik dalam mengklasifikasikan status imunisasi HB-0 pada balita di Provinsi Jawa Barat, baik untuk balita yang menerima maupun yang tidak menerima imunisasi tanpa bias terhadap salah satu kelas. Hasil analisis secara global dengan model Random Forest menunjukkan enam peubah penting yang memberikan kontribusi terbesar terhadap performa model adalah status imunisasi BCG, kepemilikan buku KIA/KMS, umur ibu, umur KRT, umur kehamilan pertama, dan klasifikasi Kab/Kota tempat tinggal. Analisis fitur penting menggunakan SHAP untuk satu amatan pertama menunjukkan bahwa status imunisasi BCG, kepemilikan buku KIA/KMS, klasifikasi Kab/Kota tempat tinggal meningkatkan kemungkinan balita mendapatkan imunisasi. Sebaliknya, jumlah anak (4), usia ibu (37 tahun), dan usia KRT (40 tahun) meningkatkan kemungkinan balita tidak mendapatkan imunisasi. Penelitian ini diharapkan dapat memberikan informasi berbasis data bagi pemerintah untuk merancang intervensi yang lebih efektif dalam meningkatkan cakupan imunisasi dan kesehatan anak di Indonesia, sekaligus mendukung pencapaian target kesehatan global.*

ABSTRACT. *Hepatitis B (HB-0) immunization in infants is an important step in preventing the transmission of hepatitis B from an early age and improving public health. This study aims to classify the HB-0 immunization status of infants in West Java Province. The method using is the Random Forest and LightGBM algorithms. The research results showed that the Random Forest model had a balanced accuracy of 0.8443, which was slightly higher than LightGBM (0.8357). This indicated that Random Forest performed better in classifying the HB-0 immunization status of infants in West Java Province, accurately distinguishing between those who received and did not receive the immunization without bias toward either class. The global analysis using the Random Forest model identified six feature importance that contributed the most to the model's performance: BCG immunization status, ownership of the KIA/KMS book, mother's age, household head's age, age at first pregnancy, and regency or city classification of residence. The feature importance analysis using SHAP for the first observation showed that BCG immunization status, ownership of the KIA/KMS book, and regency or city classification of residence increased the likelihood of infants receiving immunization. Conversely, the number of children (4), mother's age (37 years), and household head's age (40 years) increased the likelihood of infants not receiving immunization. This study is expected to provide data-driven insights for the government to design more effective interventions to improve immunization coverage and child health in Indonesia while also supporting the achievement of global health targets.*

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution-NonComercial 4.0 International License. Editorial of EULER: Department of Mathematics, Universitas Negeri Gorontalo, Jln. Prof. Dr. Ing. B. J. Habibie, Bone Bolango 96554, Indonesia.

1. Pendahuluan

Metode klasifikasi dengan *supervised learning* adalah metode yang digunakan untuk memisahkan dan mengalokasikan pengamatan ke dalam kategori yang telah ditentukan sebelumnya [1]. *Supervised learning* bekerja dengan data berlabel, artinya setiap contoh data dalam pelatihan sudah memiliki kategori atau kelas yang diketahui. Dengan demikian, model dapat belajar dari pola dalam data ini untuk mengidentifikasi karakteristik yang memisahkan setiap kelas [2]. Metode tersebut memiliki beberapa keunggulan signifikan, terutama dalam konteks pemodelan prediksi dan klasifikasi yang membutuhkan hasil akurat dan interpretasi yang jelas.

Algoritma *ensemble* seperti *Random Forest* (RF) dan *Light Gradient Boosting Machine* (*LightGBM*) telah menjadi pilihan utama dalam menyelesaikan berbagai permasalahan klasifikasi. RF sebagai

*Penulis Korespondensi.

Email : *ummulsyam08@gmail.com* (U. A. Syam)
Homepage : http://ejurnal.ung.ac.id/index.php/euler/index / E-ISSN : 2776-3706
© 2025 by the Author(s).

algoritma berbasis *bagging*, bekerja dengan cara membangun sejumlah pohon keputusan dari subset data yang diambil secara acak (*boostrap sampling*) untuk memberikan hasil yang stabil dan akurat. Metode ini sangat efektif dalam menangani *overfitting*, terutama pada dataset dengan varians tinggi [3]. RF juga memiliki kemampuan untuk menangani data yang besar dan fitur atau peubah yang saling berkorelasi tanpa risiko *overfitting* yang signifikan [4].

Berbeda dengan RF, *LightGBM* adalah algoritma berbasis *boosting* yang dirancang untuk mengurangi bias model. *LightGBM* bekerja secara iteratif dengan setiap model baru berusaha memperbaiki kesalahan prediksi model sebelumnya dengan memberikan bobot lebih besar pada data yang sulit diprediksi [5]. *LightGBM* mengadopsi pendekatan *gradient boosting* dan menggunakan teknik *histogram-based decision tree* yang membuat metode ini sangat efisien untuk data besar [6]. RF dan *LightGBM* tidak hanya *powerful* dalam klasifikasi tetapi juga menawarkan fleksibilitas untuk menangani berbagai kondisi data, termasuk data tidak seimbang.

Beberapa penelitian telah dilakukan terkait metode *Random Forest* dan *LightGBM* dalam berbagai studi kasus. Khairunnisa dkk [7] melakukan perbandingan algoritma RF dan *Double Random Forest* (DRF) dalam hal interpretabilitas model menggunakan pendekatan *association rule*. Nguyen dkk [8] mengkaji dampak *hyperparameter* pada kinerja model *LightGBM* dalam memprediksi beban listrik. Hasil penelitiannya menunjukkan bahwa pengaturan *hyperparameter* yang lebih spesifik dapat menghasilkan model yang lebih efektif dibandingkan menggunakan nilai *default*. Penelitian lainnya oleh Anggraini dkk [9] membandingkan kinerja algoritma RF, *Xgboost*, dan *LightGBM* dalam klasifikasi emosi berdasarkan komentar Reddit menggunakan dataset *GoEmotions*. Hasil analisisnya menunjukkan bahwa RF unggul dibandingkan dengan *Xgboost* dan *LightGBM* sehingga algoritma berbasis *bagging* lebih efektif untuk klasifikasi emosi berdasarkan komentar Reddit.

Selain aspek kinerja prediksi, interpretabilitas model juga menjadi perhatian penting. Interpretabilitas dalam pembelajaran mesin adalah kemampuan model untuk dipahami oleh manusia. Model yang baik bukan hanya akurat, tetapi juga harus mampu memberikan pemahaman mendalam mengenai faktor-faktor yang memengaruhi hasil prediksi [10]. Interpretabilitas model mencakup dua pendekatan utama yaitu interpretabilitas global, yang menggambarkan pengaruh fitur secara keseluruhan terhadap model, dan interpretabilitas lokal, yang menunjukkan kontribusi fitur terhadap prediksi individu [11]. Oleh karena itu, model yang baik tidak hanya memberikan hasil prediksi yang andal tetapi juga dapat digunakan untuk mendukung pengambilan keputusan berbasis data.

Dalam konteks kesehatan, Imunisasi hepatitis B (HB-0) merupakan salah satu intervensi paling penting yang wajib diberikan kepada bayi dalam waktu 24 jam setelah lahir untuk mencegah penularan virus hepatitis B dari ibu ke bayi selama proses persalinan [12]. Virus ini dapat menyebabkan kerusakan hati jangka panjang, termasuk sirosis dan kanker hati. Meskipun program imunisasi nasional telah berjalan, tantangan dalam mencapai cakupan yang merata tetap menjadi masalah utama, khususnya di wilayah seperti Provinsi Jawa Barat. Wilayah ini, dengan populasi yang sangat besar, menghadapi kesenjangan akses layanan kesehatan antara daerah perkotaan dan pedesaan [13]. Kesenjangan ini menimbulkan pertanyaan tentang faktor-faktor yang memengaruhi keputusan keluarga untuk membawa anak mereka ke fasilitas kesehatan untuk mendapatkan imunisasi HB-0.

Oleh karena itu, penelitian ini bertujuan untuk mengidentifikasi faktor-faktor yang memengaruhi penerimaan imunisasi pada balita di Provinsi Jawa Barat dan memberikan wawasan baru untuk mendukung kebijakan kesehatan. Berbeda dengan penelitian sebelumnya yang menggunakan algoritma *Random Forest* dan *LightGBM* dalam klasifikasi kesehatan, penelitian ini secara khusus mengaplikasikan kedua metode tersebut pada kasus ketimpangan cakupan imunisasi HB-0. Selain itu, pendekatan yang digunakan tidak hanya berfokus pada akurasi prediksi tetapi juga mengeksplorasi interpretabilitas model untuk memahami kontribusi setiap fitur dalam pengambilan keputusan. Evaluasi dilakukan dengan mempertimbangkan kondisi data yang tidak seimbang dan seimbang, serta membandingkan efektivitas strategi penanganan data yang umum digunakan dalam domain kesehatan. Melalui pendekatan ini, penelitian diharapkan dapat memberikan perspektif baru dalam perancangan intervensi imunisasi berbasis model prediktif yang lebih transparan dan efektif. Melalui analisis kinerja prediksi dan interpretabilitas model, penelitian ini diharapkan dapat berkontribusi pada peningkatan cakupan imunisasi dan kesehatan anak di Indonesia, serta mendukung pencapaian target global.

2. Metode

2.1. Data Penelitian

Penelitian ini menggunakan data yang bersumber dari Survei Sosial Ekonomi Nasional (SUSENAS) tahun 2023 Provinsi Jawa Barat. Jumlah amatan sebanyak 6.589 balita usia 0-5 tahun. Peubah respon yang digunakan adalah Status Imunisasi Hepatitis B (HB-0) yang kurang dari 7 hari setelah lahir. Adapun peubah penjelas yang digunakan pada penelitian ini disajikan pada Tabel 1.

2.2. Prosedur Analisis

Penelitian ini menggunakan algoritma *Random Forest* dan *LightGBM* untuk mengklasifikasikan status imunisasi hepatitis B (HB-0) pada balita di Provinsi Jawa Barat. Berikut adalah tahapan analisis yang dilakukan dalam penelitian ini:

1. Melakukan eksplorasi awal untuk memahami karakteristik data.
2. Melakukan pra-pemrosesan data, sebagai berikut:
   - Membagi dataset menjadi data latih (80%) dan data uji (20%)
   - Melakukan standarisasi peubah numerik
   - Menangani data tidak seimbang dengan beberapa metode yaitu *Random Over-Sampling* (ROS), *Random Under-Sampling* (RUS), *Synthetic Minority Over-sampling TEchnique* (SMOTE), dan *class weight*.
3. Melakukan klasifikasi menggunakan RF dan *LightGBM* dengan melakukan optimasi *hyperparameter* menggunakan *grid search* untuk mendapatkan kombinasi parameter terbaik. Model dilatih pada data seimbang dan data asli untuk membandingkan performa kedua pendekatan. Analisis dilakukan dengan mengkombinasikan *hyperparameter grid search* dan *cross-validation* dengan langkah-langkah utama sebagai berikut:
   - Step 1: *Grid search* dan *Cross Validation* (CV), kombi-

Tabel 1. Informasi peubah penjelas

Faktor	Peubah	Satuan
Karakteristik ibu	Kepemilikan buku Kesehatan Ibu dan Anak (KIA)/Kartu Menuju Sehat (KMS)	Nominal
	Umur ibu	Scale
	Pendidikan ibu	Nominal
	Status bekerja ibu	Nominal
	Akses terhadap internet	Nominal
	Kepemilikan jaminan kesehatan	Nominal
	Umur kehamilan pertama	Scale
	Tempat persalinan	Nominal
	Penolong persalinan	Nominal
Karakteristik Kepala Rumah Tangga (KRT)	Umur KRT	Scale
	Pendidikan KRT	Nominal
	Status bekerja KRT	Nominal
Karakteristik keluarga	Status Imunisasi Bacillus Calmette Guerin (BCG)	Nominal
	Status kemiskinan	Nominal
	Status kepemilikan usaha mikro kecil/menengah	Nominal
	Jumlah anak kandung	Scale
Bantuan dan subsidi	Status penerima Kartu Keluarga Sejahtera (KKS)	Nominal
	Status penerima Program Keluarga Harapan (PKH)	Nominal
	Status penerima bantuan sosial rutin	Nominal
	Status penerima bantuan sosial tidak rutin	Nominal
	Status penerima Bantuan Langsung Tunai Bahan Bakar Minyak (BLT BBM)	Nominal
	Status penerima Bantuan Langsung Tunai Desa (BLT Desa)	Nominal
	Status penerima Bantuan Padat Karya Tunai Desa (PKTD)	Nominal
Lokasi dan Wilayah	Klasifikasi Kab/Kota tempat tinggal	Nominal
	Klasifikasi Desa/Kelurahan tempat tinggal	Nominal

nasi *hyperparameter* yang dioptimasi untuk model RF yaitu *mtry*, *trees*, dan *min_n,* sedangkan untuk model *LightGBM* yaitu *mtry*, *learning rate*, *trees*, *min_n*, dan *tree_depth.* Prosedur validasi silang (*k-fold* CV) dengan nilai $k = 10$ digunakan untuk mengevaluasi kinerja model pada berbagai subset data.

- Step 2: Melatih model RF dan *LightGBM*, model dilatih menggunakan setiap kombinasi *hyperparameter* dalam *grid search*. Model dievaluasi berdasarkan nilai *balanced accuracy* untuk menilai pengaruh *hyperparameter* terhadap kinerja model.

4. Mengidentifikasi peubah penting secara global menggunakan *permutation feature importance* untuk memahami kontribusi setiap peubah terhadap prediksi.
5. Melakukan pemodelan ulang menggunakan peubah penting yang telah diidentifikasi secara global.
6. Mengevaluasi kinerja model berdasarkan nilai *balanced accuracy* dan membandingkan kinerja prediksi antara model RF dan *LightGBM* pada berbagai skenario. *Balanced accuracy* dipertimbangkan sebagai metrik evaluasi utama untuk memastikan bahwa model yang digunakan tidak hanya mengoptimalkan prediksi untuk kelas mayoritas tetapi juga memberikan perhatian yang memadai pada kelas minoritas, hal tersebut relevan dalam konteks kebijakan dan program imunisasi [14]. Berikut adalah rangkuman skenario yang dilakukan pada penelitian ini yang disajikan pada Tabel 2.
7. Melakukan pemodelan ulang menggunakan model terbaik RF dan *lightGBM*, kemudian mengevaluasi dengan 10 dataset (proporsi data latih 80% dan data uji 20%) yang diambil secara acak untuk menilai konsistensi dan stabilitas model.
8. Melakukan interpretasi terhadap model dengan performa prediksi terbaik yang terpilih secara global.
9. Melakukan interpretasi model dengan performa prediksi terbaik yang dipilih secara lokal dengan metode *SHapley Additive exPlanations* (SHAP) dengan 100 permutasi sebagai ilustrasi untuk menganalisis kontribusi setiap variabel terhadap prediksi individual. Tahapan ini hanya difokuskan pada pengamatan pertama.

Tabel 2. Informasi skenario penelitian

Skenario	Pra-proses
Skenario I (Menggunakan seluruh peubah)	- Standarisi peubah numerik tanpa penanganan data tidak seimbang - Standarisasi peubah numerik dan penanganan data tidak seimbang menggunakan ROS, RUS, SMOTE, dan *class weight*
Skenario II (Menggunakan peubah penting yang didentifikasi secara global)	- Standarisi peubah numerik tanpa penanganan data tidak seimbang - Standarisasi peubah numerik dan penanganan data tidak seimbang menggunakan ROS, RUS, SMOTE, dan *class weight*

2.3. *Random Forest*

Random Forest merupakan pengembangan dari metode *Classification and Regression Tree* (CART) yang diperkenalkan oleh Breimen tahun 2001. *Random Forest* adalah metode *ensemble learning* yang menggabungkan beberapa algoritma pohon keputusan sedemikian rupa sehingga setiap pohon dilatih dengan sampel *bootstrap* secara acak independen. Berbeda dengan prosedur pohon keputusan, *Random Forest* menambahkan pemilihan fitur acak sebelum pembentukan pohon yang disebut tahapan *random feature*

selection sehingga mengurangi korelasi antara pohon dan meningkatkan kinerja model [3]. Prediksi menggunakan *Random Forest* merupakan jumlah berbobot atas semua observasi, sehingga dapat dengan mudah direpresentasikan pada pers. (1).

$$\widehat{\mu}\left(\boldsymbol{x}\right) = \sum_{i=1}^{n} w_i(\boldsymbol{x}) Y_i \tag{1}$$

dengan w_i : rata-rata yang dihitung dengan menjumlahkan semua bobot per pohon dimana data input x jatuh ke dalam daun ℓ.

2.4. *LightGBM*

LightGBM merupakan pengembangan dari *Gradient Boosting Decision Tree* (GBDT) yang diperkenalkan oleh *Microsoft Corporation* pada tahun 2017. Kelemahan dari GBDT adalah secara komputasi membutuhkan waktu yang lebih lama ketika dihadapkan pada data dalam jumlah besar, sehingga metode *LightGBM* dikembangkan untuk meningkatkan kinerja GBDT [5]. *LightGBM* mengusulkan algoritma pohon keputusan berbasis histogram, mengadopsi teknik *Gradient-based One-Side Sampling* (GOSS) dan *Exclusive Feature Bundling* (EFB). GOSS digunakan untuk mengoptimalkan proses pembangunan pohon dengan cara mempertahankan sampel yang memiliki gradien besar dan mengurangi jumlah sampel yang memiliki gradien kecil secara acak. Teknik ini bertujuan untuk meningkatkan efisiensi tanpa mengorbankan kualitas model. Sementara EFB mengidentifikasi dan menggabungkan fitur-fitur yang eksklusif (*mutually exclusive*) dengan pendekatan berbasis graf, sehingga mengurangi jumlah fitur yang harus diproses tanpa mengorbankan akurasi model [15].

2.5. *SHapley Additive exPlanations (SHAP)*

SHAP adalah metode pasca-pemodelan yang bersifat model-agnostik dan dapat diterapkan pada model pembelajaran mesin apapun. Metode ini didasarkan pada teori permainan yang menghitung kontribusi setiap pemain terhadap hasil akhir. Dalam konteks pembelajaran mesin, pemain tersebut diwakili oleh fitur-fitur, sementara hasil akhir digantikan oleh output model. SHAP memberikan skor pada setiap fitur dalam model, yang menunjukkan tingkat pengaruhnya terhadap output model. Dalam prosesnya, SHAP mempertimbangkan berbagai kombinasi fitur, termasuk penggunaan seluruh fitur maupun subsetnya, untuk mengevaluasi semua kemungkinan kontribusi dalam model [16].

3. Hasil dan Pembahasan

3.1. *Eksplorasi Data*

Jumlah balita yang berusia 0-5 tahun dalam penelitian ini sebanyak 6.589 yang terdiri dari balita dengan status imunisasi hepatitis B (HB-0) sebanyak 3.928 (59,61%) sementara balita dengan status tidak imunisasi hepatitis B (HB-0) sebanyak 2.661 (40,39%). Hasil eksplorasi data berdasarkan faktor karakteristik ibu menunjukkan bahwa ibu yang memiliki buku KIA/KMS cenderung balitanya diimunisasi dengan persentase 80,5% (3.587 balita) dibandingkan tidak imunisasi 19,5% (868 balita). Sebaliknya, balita dari ibu yang tidak memiliki buku KIA/KMS cenderung tidak imunisasi dengan persentase 84% (1.793 balita) dibandingkan imunisasi 16% (341 balita). Ibu dengan tingkat pendidikan SD, SMP, SMA dan PT cenderung balitanya diimunisasi dengan persentase masing-masing 54,4% (935 balita); 61,5% (1.198 balita); 60,1% (1.178 balita) dan 65,8% (584 balita) dibandingkan tidak imunisasi dengan masing-masing persentase 45,6% (783 balita); 38,5% (750 balita); 39,9% (783 balita) dan 24,2% (303 balita). Ibu yang memiliki jaminan kesehatan maupun tidak memiliki juga cenderung balitanya diimunisasi dengan masing-masing persentase 61,2% (2.854 balita) dan 55,8% (1.074 balita) dibandingkan tidak imunisasi dengan masing-masing persentase 38,8% (1.809 balita) dan 44,2% (852 balita). Balita yang imunisasi maupun tidak imunisasi memiliki ibu dengan rata-rata usia kehamilan pertama yaitu 21 tahun. Balita yang lahir dengan penolong persalinan tenaga kesehatan cenderung diimunisasi dengan persentase 66,7% (1.552 balita) dibandingkan tidak imunisasi 33,3% (775 balita). Sebaliknya, balita yang lahir dengan penolong persalinan non-nakes cenderung tidak imunisasi dengan persentase 69,6% (152 balita) dibandingkan dengan imunisasi 30,4% (55 balita). Balita yang lahir di fasilitas kesehatan juga cenderung diimunisasi dengan persentase 66,8% (1.453 balita) dibandingkan tidak imunisasi 32,2% (689 balita). Sebaliknya, balita yang lahir non-faskes cenderung tidak imunisasi dengan persentase 57,9% (212 balita) dibandingkan dengan imunisasi 42,1% (154 balita).

Berdasarkan karakteristik KRT, balita yang tidak imunisasi memiliki KRT dengan rata-rata usia 38 tahun (rentang 17-77 tahun), sedangkan balita yang diimunisasi memiliki KRT dengan rata-rata usia setahun lebih muda yaitu 37 tahun (rentang 17-17 tahun). KRT dengan tingkat pendidikan SD, SMP, SMA, PT dan Pascasarjana cenderung balitanya diimunisasi dengan persentase masing-masing 56% (1.094 balita); 60,6% (912 balita); 60,8% (1.392 balita); 63,9% (483 balita) dan 57,7% (30 balita) dibandingkan tidak imunisasi dengan masing-masing persentase 44% (859 balita); 39,4% (592 balita); 39,2% (896 balita); 36,1% (273 balita) dan 42,3% (22 balita). Berdasarkan karakteristik keluarga, balita yang tidak imunisasi memiliki keluarga dengan rata-rata jumlah anak 2 per keluarga (rentang 1-11 anak), sedangkan balita yang diimunisasi juga memiliki rata-rata jumlah anak 2, namun dengan rentang lebih rendah (1-8 anak). Balita dengan status imunisasi BCG (ya) cenderung imunisasi Hepatitis B (HB-0) dengan persentase 79,9% (3.810 balita) dibandingkan tidak imunisasi 20,1% (960 balita). Sebaliknya, balita yang status imunisasi BCG (tidak) cenderung tidak imunisasi Hepatitis B (HB-0) dengan persentase 93,5% (1.701 balita) dibandingkan imunisasi 6,5% (118 balita). Keluarga dengan status miskin maupun tidak miskin juga cenderung balitanya diimunisasi dengan masing-masing persentase 52,8% (577 balita) dan 61% (3.351 balita) dibandingkan tidak imunisasi dengan masing-masing persentase 47,2% (515 balita) dan 39% (2.146 balita). Berdasarkan bantuan dan subsidi, keluarga yang menerima BLT Desa cenderung imunisasi dengan persentase 51,5% (208 balita) dibandingkan tidak imunisasi 48,5% (196 balita). Keluarga yang menerima bantuan PKTD lebih cenderung imunisasi dengan persentase 89,5% (17 balita) dibandingkan tidak imunisasi 10,5% (2 balita). Begitupun keluarga yang tidak menerima BLT Desa dan Bantuan PKTD keduanya juga cenderung imunisasi dibandingkan tidak imunisasi.

Berdasarkan lokasi dan wilayah tinggal, balita yang lokasi tinggal di kota maupun kabupaten keduanya cenderung imunisasi dengan masing-masing persentase 63,2% (1.187 balita) dan 58,2% (2.741 balita) dibandingkan tidak imunisasi 36,8% (691 balita) dan 41,8% (1.970 balita). Begitupun balita yang tinggal di wilayah kelurahan maupun desa juga cenderung imunisasi dengan

Tabel 3. Jumlah data berdasarkan kelas peubah respon menurut jenis data

Kategori/Kelas	Data Latih					Data Uji
	Standarisasi	SMOTE	ROS	RUS	*Class Weight*	
Tidak Imunisasi	2.129	3.143	3.143	2.129	2.129	532
Imunisasi	3.143	3.143	3.143	2.129	3.143	785
Total	5.272	6.286	6.286	4.258	5.272	1.317

Tabel 4. Evaluasi kinerja model *Random Forest* dan *LightGBM* dalam 2 skenario

Random Forest						
Skenario	Ukuran Evaluasi	Standarisasi	SMOTE	ROS	RUS	*Class Weight*
I	*Accuracy*	0,8664	0,8610	0,8580	0,8481	0,8641
	Sensitivity	0,9592	0,9529	0,9554	0,9121	0,9580
	Spesificity	0,7293	0,7256	0,7143	0,7538	0,7256
	BalancedAcuracy	0,8443	0,8392	0,8348	0,8329	0,8418
II	*Accuracy*	0,8368	0,8170	0,8238	0,7988	0,8375
	Sensitivity	0,9541	0,8752	0,8892	0,8395	0,9592
	Spesificity	0,6635	0,7312	0,7274	0,7387	0,6579
	BalancedAcuracy	0,8088	0,8032	0,8083	0,7891	0,8086
LightGBM						
I	*Accuracy*	0,8542	0,8565	0,8322	0,8345	0,8550
	Sensitivity	0,9427	0,9439	0,8968	0,9146	0,9452
	Spesificity	0,7237	0,7272	0,7368	0,7162	0,7218
	BalancedAcuracy	0,8332	0,8357	0,8168	0,8154	0,8335
II	*Accuracy*	0,8436	0,8360	0,8330	0,8216	0,8443
	Sensitivity	0,9720	0,9363	0,9121	0,8726	0,9732
	Spesificity	0,6541	0,6880	0,7162	0,7462	0,6541
	BalancedAcuracy	0,8131	0,8121	0,8141	0,8094	0,8137

masing-masing persentase 61% (2.698 balita) dan 56,8% (1.230 balita) dibandingkan tidak imunisasi 39% (1.726 balita) dan 43,2% (935 balita).

3.2. *Data Preprocessing*

Pada penelitian ini data dibagi menjadi dua bagian yaitu 80% untuk data latih dan 20% data uji. Peubah numerik pada data latih dilakukan proses standarisasi untuk memastikan bahwa setiap peubah memiliki skala yang sama sehingga model dapat memberikan bobot yang adil pada setiap peubah. Pada peubah respon dilakukan penanganan data tidak seimbang dengan empat metode yaitu SMOTE, ROS, RUS dan *class weight*. Konsep dari SMOTE yaitu membangkitkan data sintetis yang baru dari kelas minoritas dengan pendekatan tetangga terdekat agar diperoleh kelas yang setara dengan mayoritas [17]. ROS merupakan teknik penanganan data tidak seimbang dengan menambahkan sampel kelas minoritas secara acak hingga mencapai jumlah yang seimbang dengan kelas mayoritas, sementara RUS dengan cara menghapus sebagian sampel dari kelas mayoritas secara acak untuk menyamakan jumlahnya dengan kelas minoritas [18]. Adapun pendekatan *class weight* yaitu memberikan bobot yang lebih besar pada kelas minoritas [19]. Pada kasus ini yang menjadi kelas minoritas merupakan balita dengan tidak imunisasi hepatitis B (HB-0) sedangkan untuk kelas mayoritas yaitu balita dengan imunisasi. Formula untuk menghitung bobot kelas minoritas yaitu pembagian antara jumlah observasi kelas mayoritas dengan jumlah observasi kelas minoritas sehingga diperoleh bobotnya sebesar 1,476. Sementara itu, untuk kelas mayoritas diperoleh dengan pembagian antara jumlah observasi kelas mayoritas dengan jumlah observasi kelas mayoritas sehingga diperoleh bobotnya sebesar 1,00. Informasi jumlah data latih setelah dilakukan penanganan data tidak seimbang disajikan pada Tabel 3.

3.3. *Pemodelan Random Forest dan LightGBM*

Sebelum melakukan *fitting* model, dilakukan proses *hyperparameter tuning* yang bertujuan untuk menemukan kombinasi *hyperparameter* yang dapat menghasilkan model dengan tingkat performa yang optimal. CV digunakan untuk mengetahui kinerja model pada keseluruhan data. Pada studi ini, digunakan 10-*fold* CV yang artinya data latih dibagi menjadi 10 bagian dimana satu bagian sebagai data validasi dan sembilan bagian lainnya sebagai data latih. Berdasarkan skenario I diperoleh lima peubah penting yaitu status imunisasi BCG, kepemilikan buku KIA/KMS, umur ibu, umur KRT dan umur kehamilan pertama. Berbeda pada model *LightGBM* dengan penanganan data tidak seimbang metode ROS lima peubah pentingnya adalah status imunisasi BCG, kepemilikan buku KIA/KMS, tempat persalinan, klasifikasi Kab/Kota tempat tinggal dan status kemiskinan. Sedangkan pada model *LightGBM* dengan penanganan data tidak seimbang metode RUS lima peubah pentingnya adalah status imunisasi BCG, kepemilikan buku KIA/KMS, umur kehamilan pertama, jumlah anak, pendidikan ibu. Skenario II dibangun berdasarkan lima peubah penting yang teridentifikasi memiliki nilai *permutation-based feature importance* tertinggi dari Skenario I. Peubah-peubah ini diidentifikasi menggunakan metode *permutation feature importance*, sebuah teknik yang mengukur pentingnya suatu fitur dengan menghitung peningkatan kesalahan prediksi model setelah nilai fitur tersebut diacak [20]. Hasil evaluasi kinerja model menggunakan *Random Forest* dan *LightGBM* untuk kedua skenario disajikan pada Tabel 4.

Berdasarkan Tabel 4 model terbaik dari *Random Forest* ada-

Tabel 5. *Confusion matrix* model *Random Forest* dan *LightGBM*

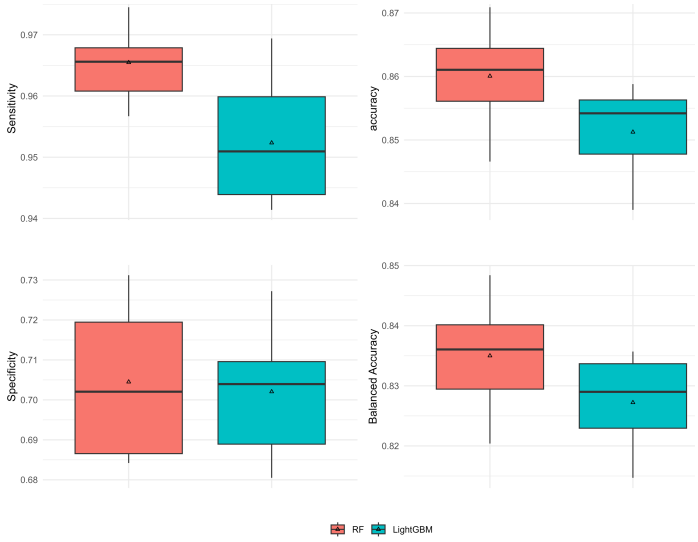
Aktual	*Random Forest*		*LightGBM*	
	Prediction		*Prediction*	
	Tidak Imunisasi	Imunisasi	Tidak Imunisasi	Imunisasi
Tidak Imunisasi	388	144	387	145
Imunisasi	32	753	44	741

lah model dengan data standarisasi tanpa penanganan data tidak seimbang pada skenario I . *Hyperparameter* optimal untuk model ini adalah jumlah pohon keputusan (*trees*) sebesar 266, jumlah prediktor yang dipilih secara acak pada setiap pemisahan (*mtry*) sebesar 18 dan ukuran minimum terminal *node* (*min_n*) sebesar 5. Nilai *balanced accuracy* pada data uji adalah 0,8443 menunjukkan bahwa secara keseluruhan performa model dalam mengklasifikasikan balita yang mendapatkan imunisasi hepatitis B (HB-0) dan yang tidak mendapatkan imunisasi sebesar 84,43%. Nilai *balanced accuracy* pada data latih dengan 10-*fold* CV sebesar 0,823 dimana nilai tersebut tidak berbeda jauh dibandingkan dengan nilai *balanced accuracy* pada data uji. Hal ini menunjukkan bahwa model memiliki kemampuan generalisasi yang baik dan tidak mengalami *overfitting*. *Accuracy* sebesar 0,8664 artinya model dapat memprediksi dengan benar 86,64% dari seluruh data. *Sensitivity* sebesar 0,9592 menujukkan bahwa model memiliki sensitivitas yang tinggi sebesar 95,92% dalam mendeteksi balita yang imunisasi. *Specificity* sebesar 0,7293 menunjukkan bahwa model dapat mengidentifikasi 72,93% balita yang tidak imunisasi dengan benar.

Model terbaik dari *LightGBM* adalah model dengan SMOTE dalam menangani data tidak seimbang pada skenario I. *Hyperparameter* optimal untuk model ini adalah nilai *trees* sebanyak 618, *mtry* adalah 20, *Min_n* sebanyak 10, kedalaman maksimum pohon (*tree_depth*) sebesar 9, dan laju pembelajaran (*learning rate*) sebesar 0.0333. Berbeda dengan *Random Forest*, pada *LightGBM* perlu diatur *tree_dept* dan *learning rate* karena *LightGBM* adalah algoritma *boosting*, dimana pohon-pohon dibuat secara berurutan untuk memperbaiki kesalahan dari pohon sebelumnya. *Learning rate* yang kecil dengan *tree_depth* yang besar dapat mengurangi *overfitting* dan mengendalikan kompleksitas model, tetapi proses pelatihan akan memerlukan lebih banyak pohon dan waktu komputasi untuk mencapai kinerja yang optimal [21]. Nilai *balanced accuracy* pada data uji menunjukkan bahwa secara keseluruhan performa model dalam mengklasifikasi balita yang mendapatkan imunisasi hepatitis B (HB-0) dan yang tidak imunisasi sebesar 83,57%. Jika dibandingkan dengan model terbaik *Random Forest*, *LightGBM* memiliki *balanced accuracy* yang lebih rendah. Nilai *accuracy* sebesar 0,8565, *sensitivity* sebesar 0,9439 dan *spesificity* sebesar 0,7272. Secara keseluruhan model ini sudah cukup baik karena nilai *balanced accuracy* pada data uji tidak berbeda jauh dengan nilai *balanced accuracy* pada data latih sebesar 0,8145. Perbandingan *confusion matrix* metode *Random Forest* dan *LightGBM* dengan model terbaik disajikan pada Tabel 5.

Tabel 5 menunjukkan *confusion matrix* dari *Random Forest* dan *LightGBM* dalam memprediksi dua kategori yaitu balita tidak imunisasi dan imunisasi. Pada *Random Forest,* model berhasil memprediksi dengan benar 388 balita yang tidak imunisasi dan 753 balita yang diimunisasi, sementara kesalahan prediksi terjadi pada 32 balita imunisasi yang diprediksi sebagai tidak imunisasi dan 144 balita tidak imunisasi yang diprediksi sebagai imunisasi. Pada *LightGBM,* model berhasil memprediksi dengan benar 387 balita yang tidak imunisasi dan 741 balita yang diimunisasi, tetapi salah memprediksi 44 balita imunisasi sebagai tidak imunisasi dan 145 balita tidak imunisasi sebagai imunisasi. Kesalahan klasifikasi pada model ini memiliki implikasi penting dalam program imunisasi. *False Positive* (FP) yaitu balita yang sebenarnya tidak menerima imunisasi diklasifikasikan sebagai telah menerima imunisasi. Kesalahan ini dapat menyebabkan *overestimate* cakupan imunisasi yang berpotensi menghambat perumusan kebijakan intervensi yang tepat sasaran. Sebaliknya, *False Negative* (FN) yaitu balita yang telah menerima imunisasi tetapi diklasifikasikan sebagai tidak diimunisasi. Hal ini dapat menyebabkan alokasi sumber daya yang tidak efisien, seperti pemberian intervensi tambahan yang sebenarnya tidak diperlukan. Oleh karena itu, model dengan tingkat FP dan FN yang rendah lebih diharapkan untuk meningkatkan akurasi pemantauan cakupan imunisasi serta efektivitas kebijakan kesehatan masyarakat.

Evaluasi performa model terbaik RF dan *LightGBM* selanjutnya dilakukan untuk menilai konsistensi dan stabilitas model. Analisis ini menggunakan *hyperparameter* terbaik dan diterapkan pada 10 dataset yang diambil secara acak. Hasil evaluasi disajikan pada Gambar 1.



Gambar 1. Evaluasi dan Performa model terbaik dengan 10 dataset secara acak

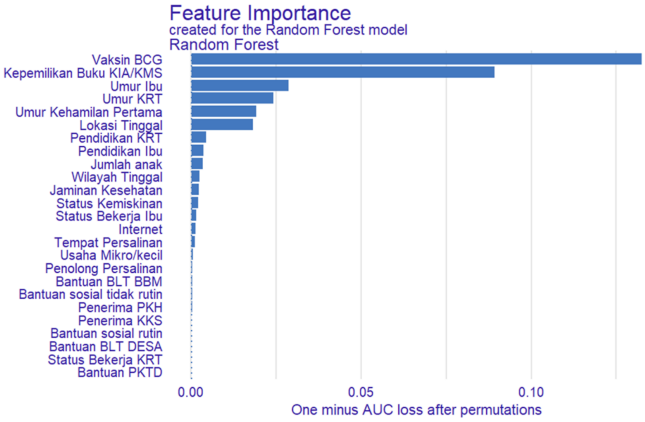
Pada analisis ini, median dipilih sebagai ukuran pemusatan karena median lebih robust terhadap outlier [22] dibandingkan rata-rata yang disimbolkan Δ pada Gambar 1. Rata-rata cenderung terpengaruh oleh nilai ekstrem sehingga dapat memengaruhi interpretasi kinerja model. Penggunaan median dalam mengevaluasi performa model dapat meningkatkan ketahanan analisis terhadap outlier, menghasilkan interpretasi yang lebih stabil, ser-

ta memberikan gambaran yang lebih representatif.

Gambar 1 menunjukkan perbandingan performa metode RF dan *LightGBM* berdasarkan empat metrik evaluasi. RF memiliki median sensitivitas dan akurasi yang lebih tinggi dibandingkan *LightGBM*. Pada spesifisitas, *LightGBM* memiliki median yang sedikit lebih tinggi dibandingkan RF. Fokus utama penelitian ini pada *balanced accuracy*, dimana RF menunjukkan median yang lebih tinggi dibandingkan *LightGBM*. Hal ini menunjukkan bahwa RF lebih andal dalam mengklasifikasikan data secara konsisten [23]. Oleh karena itu, RF dipilih sebagai model terbaik untuk mengklasifikasikan status imunisasi hepatitis B (HB-0) pada balita di Provini Jawa Barat.

3.4. *Interpretasi secara global model terbaik*

Analisis lebih lanjut dilakukan dengan memahami kontribusi masing-masing peubah secara global pada model terbaik yaitu *Random Forest* menggunakan seluruh data. Hasil analisis mengenai peubah penting dalam model ini disajikan pada Gambar 2.



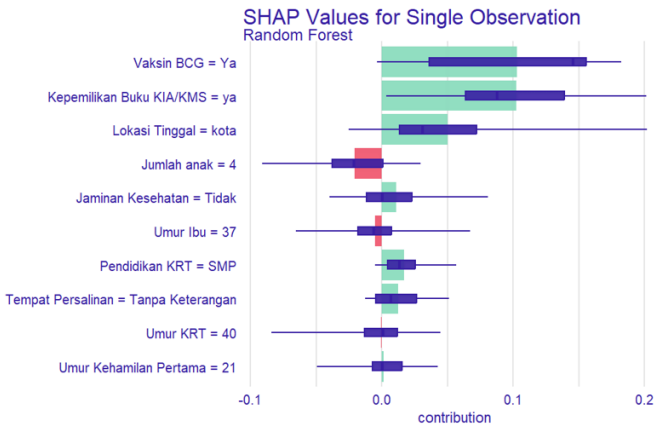
Gambar 2. Peubah penting model terbaik *Random Forest*

Gambar 2 menunjukkan kontribusi masing-masing peubah terhadap performa model RF dalam mengklasifikasikan status imunisasi hepatitis B (HB-0) pada balita di Provinsi Jawa Barat. Berdasarkan hasil analisis, enam peubah penting yang memberikan kontribusi terbesar terhadap performa model adalah imunisasi/vaksin BCG [24], kepemilikan buku KIA/KMS [25], umur Ibu [26], umur KRT [27], umur kehamilan pertama [29], dan lokasi tinggal [26].

3.5. *Ilustrasi interpretasi secara lokal dengan SHAP*

Nilai SHAP digunakan untuk mengevaluasi kontribusi masing-masing peubah terhadap prediksi status imunisasi hepatitis B (HB-0) pada balita untuk satu observasi spesifik. Analisis SHAP mengidentifikasi fitur-fitur yang memberikan kontribusi positif yaitu meningkatkan kecenderungan prediksi ke arah balita imunisasi serta fitur-fitur negatif yang mendorong prediksi ke arah balita tidak imunisasi. Hasil analisis SHAP untuk satu amatan pertama disajikan pada Gambar 3.

Sebagai ilustrasi, analisis SHAP untuk pengamatan pertama menunjukkan bahwa vaksin BCG memberikan kontribusi positif yang signifikan terhadap kemungkinan balita mendapatkan imunisasi hepatitis B (HB-0). Penerimaan imunisasi/vaksin BGG pada balita mencerminkan keterlibatan dalam program imunisasi terstruktur yang secara tidak langsung meningkatkan peluang balita



Gambar 3. SHAP *value* model terbaik *Random Forest*

untuk menerima imunisasi lanjutan. Kepemilikan buku KIA/KMS juga memberikan kontribusi positif. Buku KIA/KMS merupakan satu-satunya alat pencatatan pelayanan kesehatan ibu dan anak sejak ibu hamil, melahirkan dan selama nifas hingga bayi yang dilahirkan berusia 5 tahun, termasuk pelayanan imunisasi, gizi, tumbuh kembang anak dan Keluarga Berencana (KB) [29]. Kepemilikan buku ini juga sering dikaitkan dengan kesadaran orang tua yang lebih tinggi terhadap pentingnya imunisasi dan perkembangan kesehatan anak. Lokasi tinggal di daerah kota juga memberikan kontribusi positif. Hal ini terkait dengan kemudahan akses fasilitas kesehatan di perkotaan yang mendukung pelaksanaan imunisasi secara terjadwal dan terorganisir dibandingkan kabupaten. Sementara pendidikan KRT (SMP), umur kehamilan pertama (21 tahun), tempat persalinan (tanpa keterangan) dan jaminan kesehatan ibu (tidak) memiliki nilai kontribusi positif yang sangat kecil < 0.05, sehingga ketiga fitur tersebut memiliki peluang yang kecil dalam meningkatkan kemungkinan balita mendapatkan imunisasi. Fitur negatif yang meningkatkan kemungkinan balita tidak mendapatkan imunisasi hepatitis B (HB-0) meliputi jumlah anak (4), umur ibu (37 tahun), dan umur KRT (40 tahun). Hasil ini merupakan ilustrasi yang spesifik untuk pengamatan pertama dan tidak dapat digeneralisasi tanpa analisis lebih lanjut terhadap seluruh amatan.

4. Kesimpulan

Klasifikasi status imunisasi hepatitis B (HB-0) pada balita di Provinsi Jawa Barat dianalisis menggunakan algoritma *Random Forest* dan *LightGBM*. Hasil analisis menunjukkan model terbaik adalah *Random Forest* yang memiliki *balanced accuracy* sebesar 0,8443 sedangkan *LightGBM* sebesar 0,8357. Nilai *balanced accuracy* pada *Random Forest* sedikit lebih tinggi dibandingkan *LightGBM* sehingga analisis peubah penting secara global dan lokal dilakukan menggunakan model terbaik. Hasil analisis secara global menunjukkan enam peubah penting yang memberikan kontribusi terbesar terhadap performa model adalah status imunisasi BCG, kepemilikan buku KIA/KMS, umur ibu, umur KRT, umur kehamilan pertama, dan klasifikasi Kab/Kota tempat tinggal. Sebagai ilustrasi, analisis secara lokal untuk pengamatan pertama menunjukkan bahwa peubah positif yang meningkatkan kemungkinan balita mendapatkan imunisasi hepatitis B (HB-0) meliputi status imunisasi BCG, kepemilikan buku KIA/KMS, klasifikasi Kab/Kota tempat tinggal. Adapun fitur negatif yang meningkatkan kemungkinan

balita tidak mendapatkan imunisasi hepatitis B meliputi jumlah anak (4), umur ibu (37 tahun), dan umur KRT (40 tahun).

Kontribusi Penulis. **Ummul Auliyah Syam**: Metodologi, software, investigasi, penulisan—persiapan draf asli, visualisasi. **Irdayanti**: Metodologi, investigasi, kurasi data, analisis formal, penulisan—persiapan draf asli. **Indah Magfirrah**: Metodologi, investigasi, sumber daya, kurasi data, penulisan—persiapan draf asli, validasi. **Bagus Sartono**: Konseptualisasi, penulisan—tinjauan dan penyuntingan, supervisi. **Aulia Rizki Firdawanti**: Konseptualisasi, penulisan—tinjauan dan penyuntingan, supervisi. Semua penulis telah membaca dan menyetujui versi manuskrip yang diterbitkan.

Ucapan Terima Kasih. Para penulis menyampaikan ucapan terima kasih kepada editor serta reviewer yang telah mendukung kami dalam menyempurnakan naskah ini, dan kepada semua pihak yang telah memberikan bantaun dalam penelitian ini.

Pembiayaan. Penelitian ini tidak menerima pendanaan dari pihak eksternal.

Konflik Kepentingan. Para penulis menyatakan tidak ada konflik kepentingan yang terkait dengan artikel ini.

Referensi

[1] R. A. Johnson, and D. W. Wichern, *Applied Multivariate Statistical Analysis* (6th Edition). London: Pearson Education, 2007.

[2] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*. New York: Springer, 2009, doi: 10.1007/978-0-387-84858-7.

[3] L. Breiman, "Random Forests," *Mach Learn*, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.

[4] A. Liaw, and M. Wiener, "Classification and Regression by randomForest," *R News*, vol. 2/3, pp. 18-22, Dec. 2002. [Online]. Available: https://journal.r-project.org/articles/RN-2002-022/RN-2002-022.pdf. [Accessed: 8-Dec-2024].

[5] G. Ke, et al., "LightGBM: A Highly Efficient Gradient Boosting Decision Tree," in *Conference on Neural Information Processing Systems*, USA, 2017. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2017. [Accessed: 8-Dec-2024].

[6] N. K. Singh, and M. Nagahara, "LightGBM, SHAP, and Correlation-Matrix-Heatmap-Based Approaches for Analyzing Household Energy Data: Towards Electricity Self-Sufficient Houses," *Energies (Basel)*, vol. 17, no. 17, pp. 4518, Sep. 2024, doi: 10.3390/en17174518.

[7] A. Khairunnisa, K. A. Notodiputro, and B. Sartono, "A Comparative Study of Random Forest and Double Random Forest Models from View Points of Their Interpretability," *Scientific Journal of Informatics*, vol. 11, no. 1, pp. 207–218, Feb. 2024, doi: 10.15294/sji.v11i1.48721.

[8] K.T. Nguyen, T.N. Tran, and H.T. Nguyen, "Research on the Influence of Hyperparameters on the LightGBM Model in Load Forecasting," *Engineering, Technology & Applied Science Research*, vol. 14, no. 5, pp. 17005–17010, Oct. 2024, doi: 10.48084/etasr.8266.

[9] N. Anggraini, S. J. Putra, L. K. Wardhani, F. D. U. Arif, N. Hakiem, and I. M. Shofi, "A Comparative Analysis of Random Forest, XGBoost, and LightGBM Algorithms for Emotion Classification in Reddit Comments," *JURNAL TEKNIK INFORMATIKA*, vol. 17, no. 1, pp. 88–97, May 2024, doi: 10.15408/jti.v17i1.38651.

[10] F. Doshi-Velez, and B. Kim, "Towards A Rigorous Science of Interpretable Machine Learning," arXiv preprint arXiv:1702.08608, Feb 2017.

[11] S. Lundberg, and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," arXiv preprint arXiv:1705.07874, May 2017.

[12] World Health Organization, "Hepatitis B," 2024. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/hepatitis-b. [Accessed: 8-Dec-2024].

[13] KEMENKES RI, "Profil Kesehatan Indonesia 2022," 2023. [Online]. Available: https://kemkes.go.id/id/profil-kesehatan-indonesia-2022. [Accessed: 8-Dec-2024].

[14] K. H. Brodersen, C. S. Ong, K. E. Stephan, and J. M. Buhmann, "The balanced accuracy and its posterior distribution," in *International Conference on Pattern Recognition*, 2010, pp. 3121–3124. doi: 10.1109/ICPR.2010.764.

[15] M. Tang, Z. Peng, and H. Wu, "Fault Detection for Pitch System of Wind Turbine-Driven Doubly Fed Based on IHHO-LightGBM," *Applied Sciences*, vol. 11, no. 17, pp. 8030, Aug. 2021, doi: 10.3390/app11178030.

[16] S. M. Lundberg, P. G. Allen, and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," in *Conference on Neural Information Processing Systems*, USA, 2017. [Online]. Available: https://papers.nips.cc/paper_files/paper/2017. [Accessed: 8-Dec-2024].

[17] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, Jun. 2002, doi: 10.1613/jair.953.

[18] Haibo He, and E. A. Garcia, "Learning from Imbalanced Data," *IEEE*, vol. 21, no. 9, pp. 1263–1284, Sep. 2009, doi: 10.1109/TKDE.2008.239.

[19] M. Du, et al., "A Skew-Sensitive Evaluation Framework for Imbalanced Data Classification," *arXiv preprint*, arXiv:2010.05995, Oct. 2020.

[20] C. Molnar, G. Casalicchio, and B. Bischl, "Interpretable Machine Learning," in *Workshops of the European Conference on Machine Learning and Knowledge Discovery in Databases*, Belgium, 2020, doi: 10.1007/978-3-030-65965-3_28.

[21] T. Chen, and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, USA, 2016, doi: 10.1145/2939672.2939785.

[22] A. Agresti, and C. Franklin, *Statistics: The Art and Science of Learning from Data*, 3rd ed., Boston: Pearson, 2013.

[23] F. Duran, F. Wijaya, Y. R. Hulu, M. Harahap, and A. Prabowo, "Perbandingan Kinerja Algoritma Random Forest Classifier dan LightGBM Classifier untuk Prediksi Penyakit Jantung," *Data Sciences Indonesia*, vol. 3, no. 2, pp. 98–103, May 2024, doi: 10.47709/dsi.v3i2.3831.

[24] A. Scheid, et al., "Adjuvant effect of Bacille Calmette-Guérin on hepatitis B vaccine immunogenicity in the preterm and term newborn," *Frontiers in Immunology*, vol. 9, Jan. 2018, doi: 10.3389/fimmu.2018.00029.

[25] L. Lensiana, Y. Putri, and T. Rossita, "Hubungan Pemanfaatan Buku KIA dengan Kepatuhan Imunisasi Dasar Lengkap pada Bayi Usia 0-11 Bulan di Puskesmas Kampung Delima Kabupaten Rejang Lebong Tahun 2024," *Journal Student Scientific*, vol. 3, no. 1, pp. 77–84, Jan. 2024.

[26] S. Sukrisno, and M. Nadjib, "Analisis Faktor-Faktor yang Berhubungan dengan Akses Imunisasi Hepatitis B 0-7 Hari (HB 0) pada Bayi Baru Lahir di Indonesia Tahun 2016," *Media Kesehatan Masyarakat Indonesia*, vol. 21, no. 4, pp. 248–260, Aug. 2022, doi: 10.14710/mkmi.21.4.248-260.

[27] A. I. Rahmawati, and C. U. Wahjuni, "Faktor yang Mempengaruhi Kelengkapan Imunisasi Dasar di Kelurahan Krembangan Utara," *Jurnal Berkala Epidemiologi*, vol. 2, no. 1, pp. 59–70, Jan. 2014.

[28] M. S. Atteraya, I. H. Song, N. B. Ebrahim, S. Gnawali, E. Kim, and T. Dhakal, "Inequalities in Childhood Immunisation in South Asia," *International Journal of Environmental Research and Public Health*, vol. 20, no. 3, Feb. 2023, doi: 10.3390/ijerph20031755.

[29] KEMENKES RI, *Buku Kesehatan Ibu dan Anak (KIA)*, 2024. [Online]. Available: https://regulasi.bkpk.kemkes.go.id/detail/9d1e6348-456f-4aeb-ab0e-d3842f713f2b. [Accessed: 8-Dec-2024].