



Analisis Prediktif Faktor Kematian Balita menggunakan Logistic Regression, Random Forest, dan XGBoost

Aqila Kharismawardani, Denny Ganjar Purnama*

Fakultas Teknik dan Desain, Sistem Informasi, Universitas Pembangunan Jaya, Tangerang Selatan, Indonesia

Email: ¹aqila.kharismawardani@student.upj.ac.id, ^{2,*}Denny.ganjar@upj.ac.id

Email Penulis Korespondensi: Denny.ganjar@upj.ac.id

Abstrak—Angka Kematian Balita (AKABA) merupakan isu krusial di Indonesia yang memerlukan intervensi berbasis data. Penelitian ini bertujuan untuk mengembangkan model prediktif guna mengidentifikasi faktor-faktor risiko yang paling berpengaruh terhadap kematian balita di Kota Bandung, serta membandingkan performa tiga algoritma *machine learning*. Penelitian ini menggunakan data sekunder dari portal Open Data Kota Bandung periode 2019-2023. Metode yang digunakan adalah analisis komparatif antara Regresi Logistik, *Random Forest*, dan *XGBoost*. Untuk mengatasi masalah ketidakseimbangan kelas yang signifikan pada data, teknik *Synthetic Minority Over-sampling Technique* (SMOTE) diterapkan pada data latih. Hasil evaluasi menunjukkan bahwa ketiga model memiliki akurasi yang tinggi, namun performa pada kelas minoritas (kasus kematian) masih menjadi tantangan, yang ditunjukkan oleh F1-Score (0.12 untuk *Random Forest* dan 0.17 untuk *XGBoost*). Meskipun demikian, analisis *feature importance* dari model *Random Forest* berhasil mengidentifikasi 'penyebab_LAIN-LAIN', 'penyebab_DEMAM', dan ketersediaan tenaga kesehatan (PERAWAT, BIDAN) sebagai prediktor paling signifikan. Penelitian ini menyoroti pentingnya wawasan dari *feature importance* dalam mengidentifikasi faktor risiko pada data medis yang tidak seimbang, yang dapat memberikan landasan untuk rekomendasi kebijakan kesehatan yang lebih terarah.

Kata Kunci: Analisis Prediktif; Kematian Balita; *Machine Learning*; Data Tidak Seimbang; SMOTE

Abstract—The Under-Five Mortality Rate (UFMR) is a crucial issue in Indonesia that requires data-driven interventions. This study aims to develop a predictive model to identify the most influential risk factors for under-five mortality in Bandung City and to compare the performance of three machine learning algorithms. This research utilizes secondary data from the Bandung City Open Data portal for the period 2019-2021. The method employed is a comparative analysis of Logistic Regression, Random Forest, and XGBoost. To address the significant class imbalance in the data, the Synthetic Minority Over-sampling Technique (SMOTE) was applied to the training data. The evaluation results show that all three models achieve high accuracy, however, performance on the minority class (mortality cases) remains challenging, indicated by low F1-scores (0.12 for Random Forest and 0.17 for XGBoost). Nonetheless, the feature importance analysis from the Random Forest model successfully identified 'other causes' (penyebab_LAIN-LAIN), 'fever' (penyebab_DEMAM), and the availability of healthcare professionals (PERAWAT, BIDAN) as the most significant predictors. This study highlights the insight from feature importance in identifying risk factors in imbalanced medical data, providing a basis for more targeted health policy recommendations.

Keywords: Predictive Analysis; Under-Five Mortality; Machine Learning; Imbalanced Data; SMOTE

1. PENDAHULUAN

Angka Kematian Balita (AKABA) merupakan salah satu indikator kunci yang mencerminkan tingkat kesehatan dan keberhasilan pembangunan suatu negara (Akbar et al., 2025). Meskipun Indonesia telah mengalami kemajuan signifikan dalam menurunkan angka ini dari 26 kematian per 1.000 kelahiratan hidup pada tahun 2010 menjadi 19,38 pada tahun 2020, prioritas utama dalam agenda Kesehatan nasional masih berupa pencapaian target Sustainable Development Goals (SDGs). Kematian balita disebabkan oleh berbagai macam faktor yang kompleks, Dimana setiap faktor saling terkait satu sama lain. Adanya disparitas yang signifikan antar wilayah di Indonesia juga menunjukkan bahwa faktor risiko dan Tingkat kematian balita berbeda-beda secara geografis, yang menunjukkan bahwa perlu dilakukan analisis pada tingkat local untuk merancang intervensi yang efektif ((Dwi Putri et al., 2021)).

Penyebab langsung kematian balita seringkali diakibatkan oleh komplikasi medis saat lahir, seperti Berat Badan Lahir Rendah (BBLR), infeksi neonates, asfiksia, dan kelainan kongenital (Bitew et al., 2020). Selain itu, faktor-faktor lingkungan eksternal juga dapat mengakibatkan kematian balita, seperti sanitasi yang buruk, kualitas sarana air bersih yang tidak memadai, dan pengelolaan limbah rumah tangga, faktor tersebut memiliki kontribusi tinggi terhadap kematian balita, terutama akibat penyakit menular seperti diare ((Agus Iryanto et al., 2021)). Studi di Kota Bandung membahas pentingnya faktor BBLR dan pelayanan kesehatan seperti kunjungan neonatal dan pemberian ASI eksklusif dalam mempengaruhi kelangsungan hidup bayi (Dwi Putri et al., 2021). Adanya hubungan yang rumit antara faktor kematian balita menjadikannya sebuah tantangan yang besar, sehingga diperlukan pendekatan analitis yang mampu mengurai pola dari data yang kompleks.

Seiring dengan berkembangnya teknologi, *machine learning* telah dimanfaatkan di sektor Kesehatan yang menawarkan solusi inovatif untuk analisis data medis yang rumit dan bervolume besar (Akbar et al., 2025). Sudah terdapat banyak penelitian yang menunjukkan bahwa pendekatan statistik konvensional seperti analisis regresi memiliki keterbatasan dalam kemampuan prediktif untuk masalah mortalitas yang kompleks (Mishra et al., 2024a; Rahman et al., 2022). Penyebab langsung kematian bayi, seperti Berat Badan Lahir Rendah (BBLR), infeksi, dan asfiksia, menciptakan pola data yang rumit dan seringkali non-linear, sehingga memerlukan pendekatan analitis yang lebih canggih (Yustisi Irkan & Aril Ahri, 2022). Untuk melakukan prediksi kematian balita di Bandung, penelitian ini akan membandingkan tiga algoritma yang populer dan terbukti efektif. Logistic Regression dipilih sebagai model dasar karena efisiensinya

dalam komputasi, kemudahan interpretasi, dan telah terbukti mampu memberikan hasil akurasi yang tinggi dan konsisten dalam studi klasifikasi Kesehatan (Mfateneza et al., 2022; Mishra et al., 2024a).

Selanjutnya, Random Forest (RF), merupakan metode *ensemble learning* yang dipilih dikarenakan ketahanannya terhadap *overfitting* dan kemampuannya untuk menangani data dengan dimensi tinggi serta mengidentifikasi variabel prediktor yang paling penting (Wijaya & Rachmat, 2024a). Beberapa studi menunjukkan bahwa RF memiliki performa prediktif yang lebih unggul dibandingkan dengan metode lain dalam memprediksi mortalitas bayi (Lee et al., 2021; Mfateneza et al., 2022). Pilihan ini sejalan dengan penelitian lain yang secara konsisten menemukan RF sebagai model dengan kinerja superior untuk prediksi kematian bayi dibandingkan dengan Regresi Logistik dan metode lainnya (Rahman et al., 2022). Perbandingan dari ketiga metode ini merupakan sebuah hal yang penting untuk dilakukan, terutama karena data medis seringkali mengalami ketidakseimbangan kelas, di mana jumlah kematian jauh lebih sedikit dibandingkan dengan jumlah yang bertahan hidup, hal ini dapat menyebabkan model menjadi bias dan memiliki sensitivitas rendah (Salam et al., 2025a; Wijaya & Rachmat, 2024a). Oleh karena itu, penggunaan teknik *oversampling* seperti SMOTE terbukti efektif untuk meningkatkan performa model klasifikasi pada konteks medis yang serupa (Irfannandhy et al., 2024a; Jasman et al., 2022a).

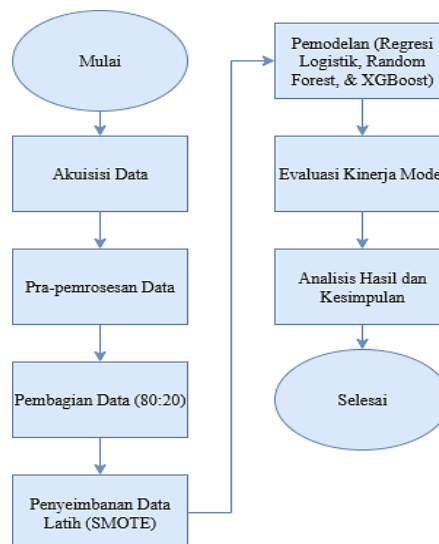
Meskipun penelitian sebelumnya telah membandingkan berbagai model *machine learning* untuk mortalitas (Saroj et al., 2022), penelitian ini secara spesifik berfokus pada perbandingan model *ensemble* (Random Forest, XGBoost) dengan model linear (Regresi Logistik) pada dataset lokal kematian balita di Indonesia (Dwi Putri et al., 2021), yang dikombinasikan dengan teknik SMOTE untuk menangani ketidakseimbangan data secara eksplisit, yang merupakan sebuah kombinasi yang masih perlu dieksplorasi lebih lanjut dalam konteks data kesehatan di Indonesia.

Berdasarkan latar belakang yang telah disusun, penelitian ini bertujuan untuk mengembangkan model prediktif untuk mengidentifikasi faktor risiko yang paling berpengaruh terhadap kematian balita di Bandung, membandingkan kinerja algoritma Logistic Regression, Random Forest, dan XGBoost dilakukan untuk menentukan model yang paling efektif, dan mengidentifikasi variabel prediktor paling signifikan dari model terbaik. Penelitian ini diharapkan dapat memberikan kontribusi yang signifikan, baik secara keilmuan melalui analisis komparatif pada data lokal, maupun secara praktis. Model prediktif yang dihasilkan dapat menjadi alat bantu bagi tenaga kesehatan dan pembuat kebijakan untuk melakukan deteksi dini terhadap balita beresiko tinggi, sehingga dapat melakukan intervensi yang lebih cepat dan tepat sasaran, yang pada akhirnya mendukung upaya percepatan penurunan Angka Kematian Balita di tingkat daerah.

2. METODOLOGI PENELITIAN

2.1 Kerangka Dasar Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan jenis penelitian analisis prediktif. Data yang digunakan merupakan data sekunder yang bersumber dari portal Open Data Kota Bandung periode 2019-2023. Lokasi penelitian secara kontekstual adalah Kota Bandung. Variabel penelitian terdiri dari variabel independent yang mencakup identitas wilayah (kecamatan, puskesmas), penyebab kematian, dan variabel kuantitatif (jumlah tenaga kesehatan, populasi balita). Variabel dependen adalah indikator kematian, sebuah variabel biner (0 atau 1) yang diturunkan dari jumlah_kematian. Teknik analisis yang digunakan dalam penelitian ini adalah perbandingan tiga algoritma klasifikasi *machine learning*: Regresi Logistik, *Random Forest*, dan *XGBoost*. Untuk mengatasi sifat data yang tidak seimbang, di mana jumlah kasus kematian jauh lebih sedikit, diterapkan teknik penyeimbangan data *Synthetic Minority Over-sampling Technique* (SMOTE). Proses penelitian ini mengikuti alur kerja ilmu data yang terstruktur, seperti diilustrasikan pada Gambar 1.



Gambar 1. Diagram alur tahapan penelitian

Tahapan penelitian yang terdapat pada Gambar 1 dapat dijelaskan sebagai berikut:

- Akuisisi dan Pra-pemrosesan Data: Data dari portal Open Data Kota Bandung dibersihkan. Variabel jumlah_kematian ditransformasi menjadi variabel target biner indikator_kematian (1 jika terjadi kematian, 0 jika tidak). Variabel kategorikal diubah menjadi format numerik menggunakan *One-Hot Encoding*, dan nilai yang hilang pada data numerik diisi menggunakan imputasi rata-rata (*mean*).
- Pembagian Data: Dataset dibagi menjadi 80% data latih dan 20% data uji.
- Penyeimbangan Data: Distribusi data kematian balita bersifat tidak seimbang (kelas minoritas). Untuk mengatasi hal tersebut, digunakan teknik *Synthetic Minority Over-sampling Technique* (SMOTE) yang diterapkan hanya pada data latih untuk membuat sampel sintesis untuk kelas minoritas.
- Pemodelan: Tiga model (*Logistic Regression*, *Random Forest*, *XGBoost*) dilatih menggunakan data latih yang telah diseimbangkan.
- Evaluasi Kinerja Model: Kinerja model dievaluasi menggunakan data uji dengan metrik Akurasi, Presisi, *Recall*, F1-Score, dan AUC (*Area Under the Curve*) yang dihitung dari *confusion matrix*.
- Analisis dan Kesimpulan: Hasil dari model terbaik dianalisis untuk mengidentifikasi faktor-faktor yang paling berpengaruh terhadap kematian balita.

2.2 Kajian Pustaka Metode

Penelitian ini membandingkan tiga algoritma klasifikasi dan satu teknik penanganan data tidak seimbang

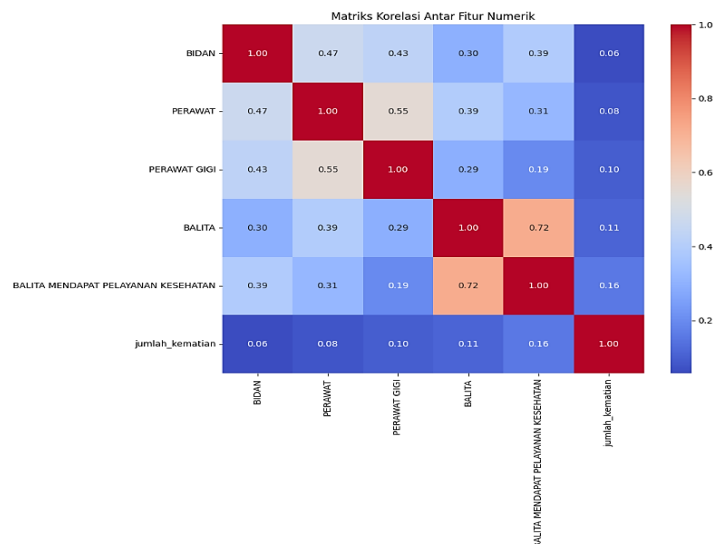
- Regresi Logistik: Merupakan model statistik linear yang digunakan untuk memprediksi probabilitas suatu kejadian biner. Model ini dipilih karena kemudahannya dalam menginterpretasikan koefisien setiap fitur untuk memahami pengaruhnya terhadap variabel target (Jasman et al., 2022b)
- Random Forest*: Merupakan sebuah metode ensemble learning yang membangun beberapa pohon keputusan (decision trees) selama pelatihan. Model ini dikenal karena kemampuannya dalam menangani interaksi kompleks antar variabel, ketahanannya terhadap overfitting, dan kemampuannya menyediakan ukuran pentingnya fitur (feature importance) (Wijaya & Rachmat, 2024b)
- XGBoost*: Merupakan implementasi canggih dari *gradient boosting* yang dikenal karena kecepatannya dan akurasi yang tinggi pada dataset terstruktur. Model ini digunakan sebagai pembanding *ensemble* yang kuat (Irfannandhy et al., 2024b)
- SMOTE (*Synthetic Minority Over-sampling Technique*): Karena data medis seringkali tidak seimbang (jumlah kasus positif/kematian jauh lebih sedikit), model dapat menjadi bias. SMOTE adalah teknik *oversampling* yang bekerja dengan cara membuat sampel sintesis baru untuk kelas minoritas berdasarkan data yang sudah ada, sehingga distribusi kelas menjadi lebih seimbang selama pelatihan (Salam et al., 2025b).

3. HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil analisis data, evaluasi kinerja model prediktif, dan pembahasan mendalam mengenai faktor-faktor yang paling berpengaruh terhadap kematian balita di Kota Bandung.

3.1 Hasil Analisis Korelasi

Analisis korelasi dilakukan untuk memahami hubungan linear antar fitur numerik dalam dataset. Hasilnya divisualisasikan pada Gambar 2.



Gambar 2. Matriks Korelasi antar fitur numerik

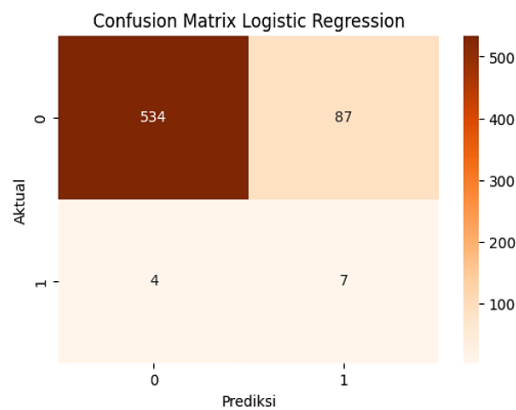
Berdasarkan Gambar 2, terdapat korelasi positif yang kuat antara populasi balita (BALITA) dan jumlah balita yang mendapat pelayanan kesehatan (BALITA MENDAPAT PELAYANAN KESEHATAN) dengan koefisien 0.72. Ini mengindikasikan bahwa seiring meningkatnya populasi balita, cakupan layanan kesehatan juga meningkat. Variabel target jumlah_kematian menunjukkan korelasi yang rendah dengan semua fitur numerik (nilai tertinggi 0.16), yang menyiratkan bahwa hubungannya kemungkinan bersifat non-linear, sehingga penggunaan model non-linear seperti *Random Forest* dan *XGBoost* menjadi relevan.

3.2 Hasil Kinerja Model

Kinerja dari ketiga model dievaluasi menggunakan data uji. Berikut adalah hasil rincinya.

3.2.1 Regresi Logistik

Evaluasi pertama dilakukan pada model Regresi Logistik. Hasil *confusion matrix* disajikan pada Gambar 3.



Gambar 3. *Confusion Matrix* Model Regresi Logistik

Sementara itu, rincian laporan klasifikasi performa model Regresi Logistik disajikan pada Tabel 1.

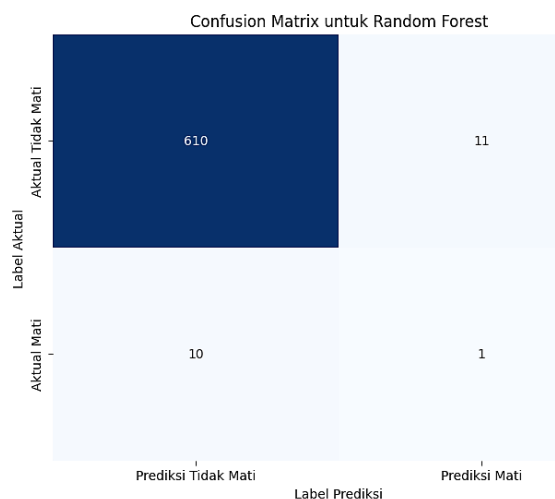
Tabel 1. Laporan Klasifikasi Model Regresi Logistik

Model	Kelas	Precision	Recall	F1-Score	Accuracy
Standar	0	0.99	0.86	0.92	0.86
	1	0.07	0.64	0.13	
Threshold Optimal (0.9869)	0	0.99	1.00	0.99	0.98
	1	0.40	0.18	0.25	

Model Regresi Logistik standar mencapai akurasi 86% dengan *recall* yang baik untuk kelas minoritas (0.64), namun presisinya sangat rendah (0.07). Setelah optimasi *threshold*, akurasi meningkat menjadi 98%, namun *recall* untuk kelas minoritas turun drastis menjadi 0.18, meskipun F1-Score sedikit membaik menjadi 0.25.

3.2.2 Random Forest

Selanjutnya, evaluasi dilakukan pada model *Random Forest*. Hasil *confusion matrix* disajikan pada Gambar 4.



Gambar 4. *Confusion Matrix* Model *Random Forest*

Rincian laporan klasifikasi untuk model *Random Forest* dapat dilihat pada Tabel 2.

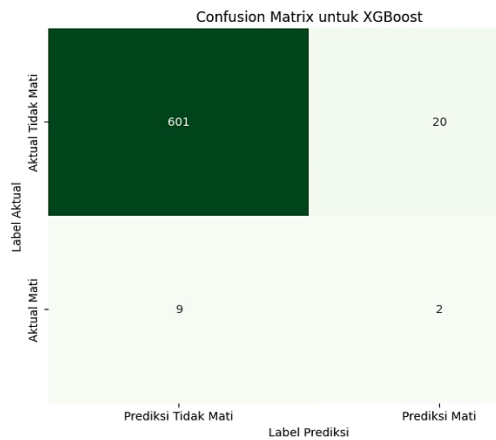
Tabel 2. Laporan Klasifikasi Model *Random Forest*

Model	Kelas	Precision	Recall	F1-Score	Accuracy
Standar	0	0.98	0.98	0.98	0.97
	1	0.08	0.09	0.09	
Threshold Optimal (0.0300)	0	0.99	0.89	0.94	0.88
	1	0.07	0.45	0.12	

Model *Random Forest* standar memiliki akurasi 97%, tetapi performanya sangat buruk pada kelas minoritas (F1-Score 0.09). Dengan *threshold* optimal, performa pada kelas minoritas membaik signifikan (*recall* 0.45), meskipun F1-Score hanya mencapai 0.12.

3.2.3 XGBoost

Evaluasi terakhir dilakukan pada model XGBoost. Hasil *confusion matrix* disajikan pada Gambar 5.



Gambar 5. *Confusion Matrix* Model XGBoost

Adapun laporan klasifikasi lengkap untuk model XGBoost disajikan pada Tabel 3.

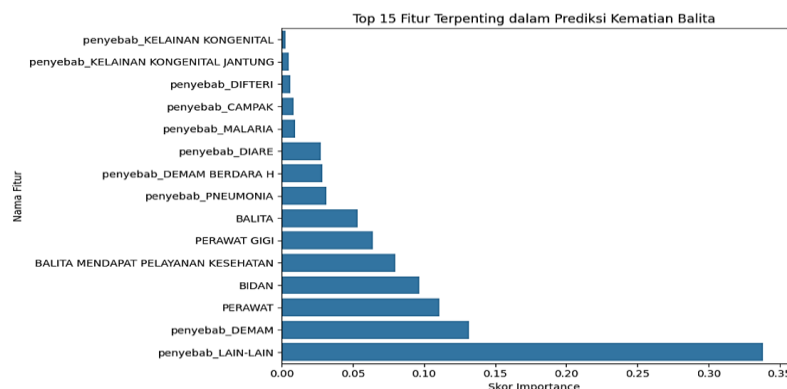
Tabel 3. Laporan Klasifikasi Model XGBoost

Model	Kelas	Precision	Recall	F1-Score	Accuracy
Standar	0	0.99	0.97	0.98	0.95
	1	0.09	0.18	0.12	
Threshold Optimal (0.9497)	0	0.98	1.00	0.99	0.98
	1	1.00	0.09	0.17	

Model XGBoost standar mencapai akurasi 95% dengan F1-Score 0.12 untuk kelas minoritas. Setelah optimasi *threshold*, model ini mencapai F1-Score tertinggi untuk kelas minoritas yaitu 0.17, dengan presisi sempurna (1.00) namun *recall* yang rendah (0.09).

3.3 Analisis Fitur Penting

Analisis ini berfokus pada hasil dari model *Random Forest* untuk mengukur kontribusi relatif setiap fitur. Hasilnya disajikan pada Gambar 6.



Gambar 6. Tingkat kepentingan 15 fitur teratas dari model *Random Forest*



Fitur yang paling berpengaruh adalah penyebab_LAIN-LAIN (skor ~0.34), diikuti oleh penyebab_DEMAM (skor ~0.13), dan serangkaian fitur terkait ketersediaan tenaga kesehatan seperti PERAWAT (skor ~0.11) dan BIDAN (skor ~0.10).

3.4 Pembahasan

Hasil penelitian menunjukkan bahwa meskipun semua model mencapai akurasi tinggi (di atas 86%), metrik ini menyesatkan karena ketidakseimbangan data. F1-Score untuk kelas minoritas (kasus kematian) menjadi tolak ukur yang lebih baik, di mana *XGBoost* dengan *threshold* optimal menunjukkan kinerja terbaik (0.17). Keunggulan model berbasis *ensemble* seperti *Random Forest* dan *XGBoost* dalam menangkap hubungan non-linear yang kompleks terbukti relevan, sejalan dengan temuan dari penelitian sebelumnya (Mishra et al., 2024b).

Analisis fitur penting menggarisbawahi bahwa penyebab_LAIN-LAIN menjadi prediktor utama. Hal ini mengindikasikan adanya penyebab kematian signifikan yang belum terkategori secara spesifik dalam dataset, sehingga memerlukan investigasi lebih lanjut oleh dinas kesehatan setempat. Peran signifikan penyebab_DEMAM dan penyebab_PNEUMONIA sejalan dengan literatur kesehatan masyarakat yang menempatkan penyakit infeksi sebagai penyebab utama mortalitas balita. Pentingnya fitur jumlah PERAWAT dan BIDAN juga menegaskan bahwa akses terhadap tenaga kesehatan profesional merupakan faktor determinan dalam kelangsungan hidup balita, sebuah temuan yang mendukung penelitian lain mengenai pentingnya infrastruktur kesehatan (Tirsa Lengkong et al., 2020). Temuan ini mendukung tujuan penelitian dalam mengidentifikasi faktor risiko utama dan menunjukkan bahwa intervensi kebijakan perlu difokuskan pada peningkatan akurasi pencatatan data kematian serta alokasi sumber daya manusia di fasilitas kesehatan.

Temuan penelitian ini, yang menunjukkan keunggulan performa *Random Forest*, konsisten dengan berbagai studi sebelumnya dalam domain prediksi mortalitas bayi. Sebagai contoh, sebuah studi juga menemukan bahwa model *Random Forest* memiliki kinerja superior (AUC 0.93) dibandingkan Regresi Logistik (AUC 0.84) dalam memprediksi risiko kematian pada bayi prematur (Lee et al., 2021). Demikian pula, penelitian yang mengilustrasikan kasus kematian bayi di Bangladesh, menyimpulkan bahwa teknik *Random Forest* menunjukkan kinerja prediktif yang lebih tinggi dibandingkan pendekatan lain, termasuk Regresi Logistik (Rahman et al., 2022). Di Rwanda, penelitian lain juga mengidentifikasi *Random Forest* sebagai klasifikasi terbaik dengan akurasi mencapai 84.3% (Mfateneza et al., 2022). Penggunaan teknik SMOTE untuk mengatasi data tidak seimbang dalam penelitian ini juga didukung oleh temuan penelitian lain, yang menyatakan bahwa kombinasi algoritma *Random Forest* dengan SMOTE cocok digunakan untuk klasifikasi data medis yang tidak seimbang, seperti pada kasus prediksi gagal jantung (Jasman et al., 2022a). Hal ini memperkuat validitas pendekatan metodologis yang digunakan dan menunjukkan bahwa kombinasi *Random Forest* dengan SMOTE merupakan strategi yang andal untuk kasus prediksi medis dengan data minoritas.

4. KESIMPULAN

Penelitian ini berhasil membandingkan model Regresi Logistik, *Random Forest*, dan *XGBoost* untuk memprediksi faktor kematian balita di Bandung menggunakan teknik penyeimbangan data SMOTE. Hasil evaluasi menunjukkan bahwa meskipun model *ensemble* (*Random Forest* dan *XGBoost*) menunjukkan potensi kinerja prediktif yang lebih baik dalam menangani data kompleks, tantangan utama tetap ada dalam memprediksi kelas minoritas (kasus kematian) secara akurat, yang dibuktikan dengan nilai F1-Score tertinggi hanya mencapai 0.17 (*XGBoost*). Meskipun kinerja prediktif model masih rendah untuk kasus minoritas, kontribusi utama penelitian ini terletak pada hasil analisis *feature importance* yang berhasil mengidentifikasi faktor-faktor paling signifikan, yaitu 'penyebab_LAIN-LAIN', 'penyebab_DEMAM', dan ketersediaan tenaga kesehatan (PERAWAT, BIDAN). Temuan ini membuktikan efektivitas kombinasi *machine learning* dan SMOTE dalam mengekstraksi wawasan (bukan hanya prediksi) dari data medis yang tidak seimbang, yang dapat menjadi dasar rekomendasi kebijakan. Keterbatasan utama penelitian ini adalah dominansi kategori penyebab_LAIN-LAIN yang tidak memiliki makna analitis jelas. Fitur ini justru menjadi prediktor terkuat, yang menunjukkan bahwa kualitas pencatatan data menjadi masalah utama. Oleh karena itu, penelitian selanjutnya disarankan untuk fokus pada dua area: (1) melakukan analisis lebih mendalam pada dataset (bekerja sama dengan ahli domain) untuk mengurai dan merekayasa fitur dari kategori penyebab_LAIN-LAIN tersebut, dan (2) mengeksplorasi teknik *sampling* alternatif selain SMOTE (misalnya, ADASYN atau *under-sampling*) yang mungkin lebih cocok untuk struktur data ini. Perbaikan ini krusial untuk meningkatkan akurasi dan daya guna model prediktif di masa depan.

REFERENCES

- Agus Iryanto, A., Joko, T., & Raharjo, M. (2021). Literature Review : Faktor Risiko Kejadian Diare Pada Balita Di Indonesia Literature Review : Risk Factors For The Incidence of Diarrhea in Children Under Five in Indonesia. *Jurnal Kesehatan Lingkungan*, 11(1), 1–7. <https://doi.org/10.47718/jkl.v10i2.1166>
- Akbar, I., Supriadi, F., & Junaedi, D. I. (2025). Pemanfaatan Machine Learning Di Bidang Kesehatan. In *Jurnal Mahasiswa Teknik Informatika* (Vol. 9, Issue 1).
- Bitew, F. H., Nyarko, S. H., Potter, L., & Sparks, C. S. (2020). Machine learning approach for predicting under-five mortality determinants in Ethiopia: evidence from the 2016 Ethiopian Demographic and Health Survey. *Genus*, 76(1). <https://doi.org/10.1186/s41118-020-00106-2>



- Dwi Putri, A., Devianto, D., & Yanuar, F. (2021). Pemodelan Jumlah Kematian Bayi Di Kota Bandung Dengan Menggunakan Regresi Zero-Inflated Poisson. *Jurnal Matematika UNAND*, 10(4), 464–475.
- Irfannandhy, R., Handoko, L. B., & Ariyanto, N. (2024a). Analisis Performa Model Random Forest dan CatBoost dengan Teknik SMOTE dalam Prediksi Risiko Diabetes. *Edumatic: Jurnal Pendidikan Informatika*, 8(2), 714–723. <https://doi.org/10.29408/edumatic.v8i2.27990>
- Jasman, T. Z., Hasmin, E., Sunardi, Susanto, C., & Musu, W. (2022a). Perbandingan Logistic Regression, Random Forest, dan Perceptron pada Klasifikasi Pasien Gagal Jantung. *CSRID (Computer Science Research and Its Development Journal)*, 14(3), 271–286. <https://doi.org/10.22303/csrid.14.3.2022.271-286>
- Lee, J., Cai, J., Li, F., & Vesoulis, Z. A. (2021). Predicting mortality risk for preterm infants using random forest. *Scientific Reports*, 11(1). <https://doi.org/10.1038/s41598-021-86748-4>
- Mfateneza, E., Rutayisire, P. C., Biracyaza, E., Musafiri, S., & Mpabuka, W. G. (2022). Application of machine learning methods for predicting infant mortality in Rwanda: analysis of Rwanda demographic health survey 2014–15 dataset. *BMC Pregnancy and Childbirth*, 22(1). <https://doi.org/10.1186/s12884-022-04699-8>
- Mishra, A., Vasishtha, G., & Maiti, S. (2024a). Predicting factors associated with under-5 mortality in India using machine learning algorithms: evidence from National Family Health Survey, 2019-21. <https://doi.org/10.21203/rs.3.rs-5309131/v1>
- Rahman, A., Hossain, Z., Kabir, E., & Rois, R. (2022). An assessment of random forest technique using simulation study: illustration with infant mortality in Bangladesh. *Health Information Science and Systems*, 10(1). <https://doi.org/10.1007/s13755-022-00180-0>
- Salam, A., Azhari, L., Septarini, R. S., & Heriyani, N. (2025a). Pendekatan Hybrid K-Means SMOTE dan Logistic Regression Untuk Deteksi Dini Diabetes Mellitus Pada Imbalanced Data. *Bulletin Of Computer Science Research*, 5(3), 219–227. <https://doi.org/10.47065/bulletincsr.v5i3.502>
- Saroj, R. K., Yadav, P. K., Singh, R., & Chilyabanyama, O. N. (2022). Machine Learning Algorithms for understanding the determinants of under-five Mortality. *BioData Mining*, 15(1). <https://doi.org/10.1186/s13040-022-00308-8>
- Tirsa Lengkong. (2020). Faktor-Faktor Yang Berhubungan Dengan Kematian Bayi Di Indonesia. In *Jurnal KESMAS*, 9(4).
- Wijaya, V., & Rachmat, N. (2024a). Comparison of SVM, Random Forest, and Logistic Regression Performance n Student Mental Health Screening. *JEECS (Journal of Electrical Engineering and Computer Sciences)*, 9(2), 173–184. <https://doi.org/10.54732/jeeecs.v9i2.9>
- Yustisi Irkan, N., & Aril Ahri, R. (2022). Analysis of Factors Associated with Infant Mortality. *Journal of Muslim Community Health (JMCH) 2022*, 3(1), 24–32. <https://doi.org/10.52103/jmch.v3i1.783>