

Pendekatan Explainable Machine Learning untuk Klasifikasi Tingkat Kemiskinan di Indonesia

Explainable Machine Learning Approach to Classifying Poverty Levels Across Indonesia

Muhammad Nur Arafah¹; Naila Anbar Afifah²; Margareta Alinsa Ejung³; Maria Delviana Berti⁴; Risna Handayani⁵; Rahmat Fuadi Syam^{6*} Imran Iskandar⁷

¹ Irmex Digital Akademika, Makassar 90551, Indonesia

^{2,3,4,5,6,7} Universitas Pancasakti, Jl. Andi Mangerangi No. 73, Makassar 90211, Indonesia

¹mnurarafah18@gmail.com; ²nailaanbar0@gmail.com; ³margarethaalinsaejung@gmail.com; ⁴delvimaria477@gmail.com;

⁵risnah643@gmail.com; ⁶rahmat@unpati.ac.id; ⁷imran.ikandar@unpacti.ac.id

* Corresponding author

Abstrak

Kemiskinan di Indonesia merupakan permasalahan sosial-ekonomi yang kompleks dan multidimensional, dipengaruhi oleh faktor pendidikan, pengeluaran, dan akses terhadap sumber daya dasar. Penelitian ini bertujuan untuk mengklasifikasikan tingkat kemiskinan di Indonesia secara komprehensif dengan memanfaatkan pendekatan Machine Learning berbasis algoritma Random Forest yang dikombinasikan dengan metode interpretabilitas SHAP (SHapley Additive exPlanations). Data penelitian diperoleh dari Badan Pusat Statistik (BPS) periode 2023–2024, mencakup 514 catatan provinsi dan kabupaten/kota dengan 12 variabel sosial ekonomi seperti persentase penduduk miskin, rata-rata lama sekolah, pengeluaran per kapita, Indeks Pembangunan Manusia (IPM), akses air bersih dan sanitasi, serta tingkat pengangguran terbuka. Data dibagi menjadi 70% untuk pelatihan dan 30% untuk pengujian. Hasil evaluasi menunjukkan bahwa model Random Forest Classifier mencapai akurasi sebesar 96,13%, presisi 96,29%, recall 96,13%, dan F1-score 96,17%, menunjukkan kinerja yang sangat baik dalam mengklasifikasikan kemiskinan menjadi lima kelas. Analisis interpretabilitas menggunakan SHAP mengidentifikasi bahwa fitur paling berpengaruh terhadap tingkat kemiskinan adalah persentase penduduk miskin (P0), pengeluaran per kapita, dan rata-rata lama sekolah. Peningkatan pendidikan dan pengeluaran per kapita menurunkan risiko kemiskinan, sedangkan tingginya nilai P0 dan pengangguran meningkatkan kemungkinan suatu wilayah tergolong miskin. Hasil penelitian ini tidak hanya menghasilkan model prediksi yang akurat, tetapi juga memberikan wawasan yang transparan dan dapat ditindaklanjuti bagi pembuat kebijakan dalam merancang strategi mengatasi tingkat kemiskinan yang berbasis data, berkelanjutan, dan tepat sasaran di Indonesia.

Kata Kunci: Klasifikasi, Pembelajaran Mesin, XAI, Data Mining, SHAP

Abstract

Poverty in Indonesia is a complex and multidimensional socio-economic problem, influenced by factors such as education, expenditure, and access to basic resources. This study aims to comprehensively classify the poverty level in Indonesia by utilizing a Machine Learning approach based on the Random Forest algorithm combined with the SHAP (SHapley Additive exPlanations) interpretability method. Research data obtained from the Central Statistics Agency (BPS) for the 2023–2024 period, includes 514 provincial and regency/city records with 12 socio-economic variables such as the percentage of the poor population, average length of schooling, per capita expenditure, Human Development Index (HDI), access to clean water and sanitation, and open unemployment rate. The data is divided into 70% for training and 30% for testing. The results of the evaluation showed that the Random Forest Classifier model achieved an accuracy of 96.13%, precision of 96.29%, recall of 96.13%, and an F1-score of 96.17%, showing excellent performance in classifying poverty into five classes. The interpretability analysis using SHAP identified that the most influential features on the poverty rate were the percentage of the poor population (P0), per capita expenditure, and average length of schooling. Increased education and per capita expenditure reduce the risk of poverty, while high P0 values and unemployment increase the likelihood of a region being classified as poor. The results of this study not only produce accurate prediction models, but also provide transparent and actionable insights for policymakers in designing data-based, sustainable, and targeted poverty alleviation strategies in Indonesia.

Keywords: Classification, Machine Learning, XAI, Data Mining, SHAP

Pendahuluan

Kemiskinan masih menjadi masalah sosial ekonomi yang kompleks di Indonesia, ditandai dengan kesenjangan pendapatan, pendidikan, dan akses sumber daya antar daerah [1], [2]. Ukuran kemiskinan yang hanya berfokus pada pendapatan seringkali gagal mewakili dimensi multidimensi kesejahteraan masyarakat [3], [4]. Oleh karena itu, pendekatan berbasis machine learning memberikan peluang baru untuk menganalisis dan mengklasifikasikan tingkat kemiskinan secara komprehensif berdasarkan berbagai indikator sosial ekonomi [5].

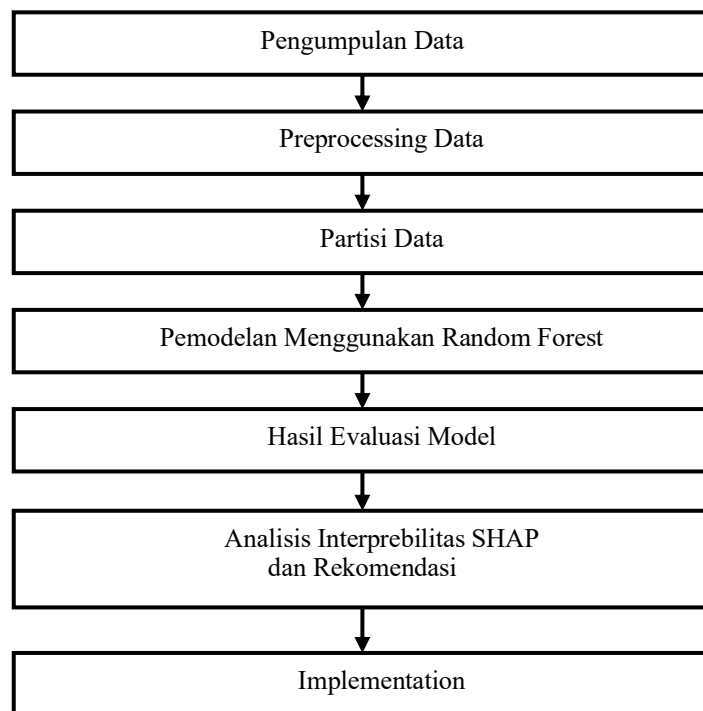
Dalam penelitian ini, data kemiskinan dari seluruh provinsi dan kabupaten/kota di Indonesia digunakan, yang mencakup indikator-indikator seperti persentase penduduk miskin, rata-rata lama sekolah, pengeluaran per kapita, indeks pembangunan manusia, harapan hidup, akses ke sanitasi dan air minum yang layak, tingkat pengangguran terbuka, partisipasi angkatan kerja, dan produk domestik regional bruto (PDB) [6], [7]. Variabel ini digunakan untuk mengklasifikasikan kategori kemiskinan daerah menggunakan algoritma Random Forest, yang diketahui memiliki kemampuan tinggi dalam menangani data dan variabel skala besar yang berinteraksi satu sama lain [8], [9].

Metode Random Forest bekerja dengan membentuk sejumlah pohon keputusan yang digabungkan menggunakan teknik bagging, sehingga meningkatkan stabilitas dan akurasi model [6], [10]. Penelitian sebelumnya telah menunjukkan bahwa Random Forest efektif dalam memetakan kemiskinan berdasarkan data spasial dan indikator sosial ekonomi [11], [12]. Selain itu, metode ini mampu memberikan perkiraan pentingnya fitur yang menunjukkan seberapa besar setiap indikator berkontribusi terhadap hasil klasifikasi [7], [13]. Ke depannya, pengembangan model ini akan diintegrasikan dengan Explainable Artificial Intelligence (XAI) seperti SHAP untuk memberikan interpretasi hasil klasifikasi [14]. Pendekatan XAI akan memungkinkan pembuat kebijakan untuk secara transparan memahami bagaimana faktor-faktor seperti pendidikan, pengeluaran, dan akses ke sanitasi memengaruhi tingkat kemiskinan di setiap wilayah [15]. Dengan demikian, penelitian ini bertujuan untuk mengembangkan model klasifikasi tingkat kemiskinan di Indonesia berdasarkan Random Forest, yang tidak hanya akurat tetapi juga berpotensi untuk dijelaskan dan ditafsirkan sebagai dasar perumusan kebijakan pembangunan berkelanjutan [4], [16].

Metode

A. Jenis Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimental untuk mengklasifikasikan tingkat kemiskinan di Indonesia.



Tabel 1. Tahapan Penelitian

B. Pengumpulan Data

Data diperoleh dari Badan Pusat Statistik (BPS) dan dianalisis menggunakan Random Forest Classifier, yang mampu mengenali pola kompleks di antara variabel sosial ekonomi dan mengidentifikasi faktor-faktor paling berpengaruh yang berkontribusi terhadap kemiskinan yang dimana data tersebut bersifat aktual sesuai dengan kondisi yang terjadi.

C. Instrumen Penelitian

Model penelitian dikembangkan menggunakan algoritma Random Forest *Ensemble* dengan Python 3.12 dan perpustakaan utama seperti NumPy, Pandas, Matplotlib, dan Scikit-learn untuk pelatihan dan evaluasi model. Analisis

interpretabilitas dilakukan dengan menggunakan SHAP (SHapley Additive exPlanations) untuk menjelaskan pengaruh masing-masing variabel terhadap hasil klasifikasi. Antarmuka Streamlit digunakan untuk pengujian data waktu nyata, dan semua eksperimen dilakukan pada komputer pribadi tanpa akselerasi GPU.

D. Preprocessing Data

Tahapan preprocessing data dilakukan untuk memastikan kualitas dan konsistensi dataset sebelum digunakan dalam pelatihan model. Proses ini mencakup pembersihan data (data cleaning) seperti penanganan nilai hilang (missing values) dan deteksi data anomali, normalisasi skala variabel agar setiap indikator memiliki bobot yang seimbang, serta transformasi data ke dalam format numerik yang sesuai dengan kebutuhan algoritma Random Forest. Selain itu, data dibagi menjadi dua bagian, yaitu data pelatihan (training data) dan data pengujian (testing data) dengan proporsi 70:30 untuk memastikan model dapat dievaluasi secara objektif terhadap data yang belum pernah dilihat sebelumnya.

E. Penerapan Metode SHAP

Selanjutnya, untuk menganalisis interpretabilitas model, digunakan metode SHAP (SHapley Additive exPlanations) yang memungkinkan identifikasi kontribusi masing-masing variabel terhadap hasil klasifikasi tingkat kemiskinan. Melalui analisis SHAP, dapat diketahui variabel mana yang memiliki pengaruh paling besar terhadap prediksi model, serta arah pengaruhnya terhadap peningkatan atau penurunan tingkat kemiskinan. Pendekatan ini membantu memberikan pemahaman yang lebih transparan mengenai keputusan model dan mendukung validasi hasil dari sisi kebijakan sosial ekonomi.

F. Evaluasi Model

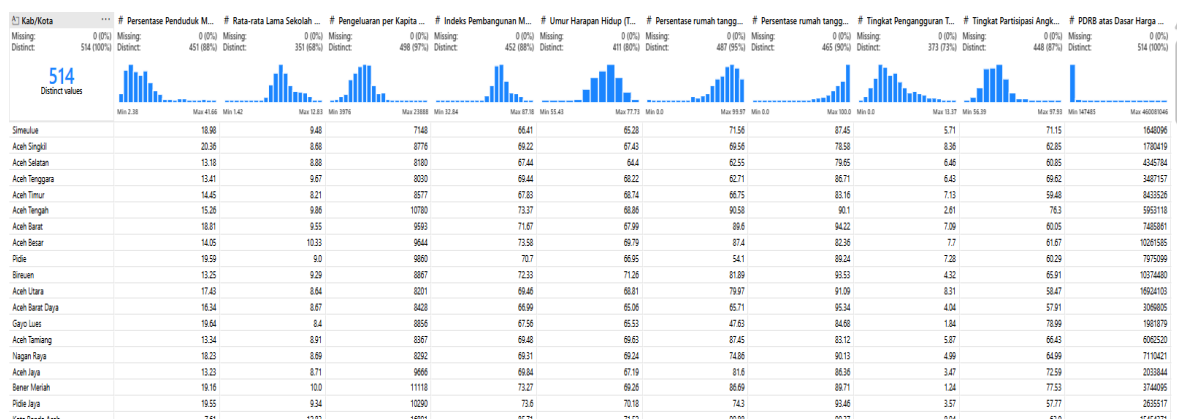
Evaluasi model Machine Learning dilakukan untuk menilai tingkat kinerja dan efektivitas model yang dikembangkan. Hasil evaluasi ditampilkan dalam bentuk *heatmap*, tabel, serta perhitungan matematis berdasarkan hasil pengujian model yang telah dilatih. Proses evaluasi menggunakan *confusion matrix* guna memperoleh informasi mengenai nilai akurasi, presisi, *recall*, dan *F1-score*. Melalui metrik ini, performa model dapat dianalisis secara menyeluruh untuk menentukan seberapa baik algoritma *Random Forest* dalam mengklasifikasikan tingkat kemiskinan pada setiap provinsi, kabupaten, dan kota di Indonesia.

Hasil dan Diskusi

Data penelitian diperoleh dari Badan Pusat Statistik (BPS) Republik Indonesia untuk tahun 2023 hingga 2024, mencakup indikator sosial ekonomi di seluruh provinsi dan kabupaten/kota.

A. Pengumpulan Data

Sebanyak 514 catatan digunakan dalam penelitian ini, dibagi menjadi 70% data pelatihan dan 30% data pengujian. Kumpulan data mencakup 12 variabel: Provinsi, Kota, Persentase Kemiskinan, Sekolah, Pengeluaran, IHD (IPM), UHH, Sanitasi, Akses Air, TPT, TPAK, dan PDRB.



Gambar 1. Perwakilan Data

B. Preprocessing Data

Sebelum pelatihan model, beberapa langkah pra-pemrosesan dilakukan pada dataset Klasifikasi Tingkat Kemiskinan di Indonesia. Kolom dengan nama panjang disederhanakan, atribut kategoris dikodekan menggunakan Pengkodean Label dan Pengkodean Satu-Panas, dan fitur berkelanjutan dinormalisasi dengan Scaler Min-Max untuk memastikan nilai berada dalam kisaran [0,1]. Seleksi fitur dilakukan dengan menggunakan Random Forest untuk mengidentifikasi variabel yang paling relevan, dan ketidakseimbangan kelas diatasi dengan menggunakan metode SMOTE.

Tabel 1. Preprocessing Data

<i>Preprocessing</i>	<i>Hasil</i>
Data Cleaning	Menangani nilai yang hilang, konversi tipe data, menangani duplikat, pembersihan data teks.
Implementation of Smote	Sebelum di SMOTE : 55, 182, 150, 61, 66 Setelah di SMOTE : 182, 182, 182, 182, 182
Target Class	Target Kelas 0-4
Feature Engineering Normalize Data	Rasio Pengeluaran PDRB ditambahkan. Normalisasi (MinMaxScaler) diterapkan pada 12 kolom numerik.

C. Partisi Data

Kumpulan data dibagi menjadi 70% data pelatihan dan 30% pengujian, dan model Random Forest Classifier dilatih untuk mencapai hasil prediksi yang akurat dan stabil.

Tabel 2. Partisi Data

Data Train	Data Test
360	154

D. Pemodelan Menggunakan Random Forest

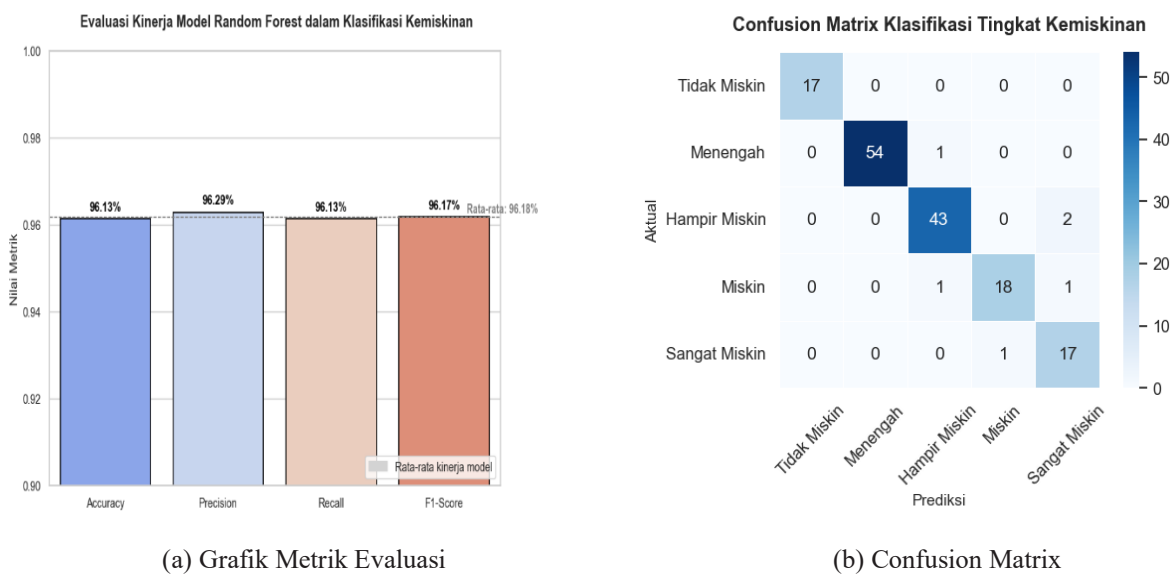
<i>Parameter</i>	<i>Fungsi</i>	<i>Justifikasi</i>
n_estimators (300)	Menentukan jumlah pohon keputusan dalam hutan.	Dari hasil tuning dengan nilai {100, 200, 300, 400}, performa validasi meningkat hingga 300, tetapi tidak berubah signifikan setelah itu (performance plateau).Jumlah 300 memberikan stabilitas prediksi dan mengurangi varians.
max_depth (12)	Mencegah agar tidak terjadi overfitting data dan Membatasi kedalaman maksimum setiap pohon, dan Membatasi kedalaman mencegah pohon belajar pola noise atau outlier	Pada tuning, kedalaman >12 menyebabkan akurasi train mendekati 100% (indikasi overfitting).Kedalaman <10 menyebabkan underfitting. Nilai 12 memberi bias-variens trade-off yang paling seimbang.
Min_samples_split (4) Min_samples_leaf (2)	diterapkan agar setiap simpul memiliki jumlah minimum sampel yang memadai sebelum dilakukan pemisahan atau sebelum menjadi simpul daun	Nilai split kecil (2–3) membuat node terlalu sensitif terhadap noise, sedangkan nilai 4 lebih stabil dan memberikan generalisasi lebih baik karena hanya membelah ketika sampel cukup akurat. Leaf size terlalu kecil (1) mudah menyebabkan overfitting, sementara leaf size 3 menurunkan akurasi. Leaf size 2 menjadi pilihan paling seimbang antara akurasi dan kompleksitas model.
Random State (42)	memastikan hasil pelatihan dapat direproduksi dengan akurat.	Pemilihan nilai 42 adalah praktik umum untuk memastikan konsistensi eksperimen. Tidak langsung mencegah overfitting, tetapi memastikan eksperimen stabil dan evaluasi dapat diverifikasi.
Class_weight (Balanced)	Sehingga model memberikan bobot yang lebih besar pada kelas-kelas dengan jumlah sampel yang lebih sedikit	Balanced weight mencegah kelas minoritas diabaikan Tanpa class_weight, model cenderung overfit pada kelas mayoritas. Balanced weight menjaga kontribusi semua kelas secara proporsional.
N_jobs (1)	memungkinkan pemrosesan paralel untuk mempercepat pelatihan model.	Mempercepat training, memungkinkan tuning lebih baik

Dalam penelitian ini, model Random Forest diterapkan untuk mengklasifikasikan tingkat kemiskinan menjadi lima kelas. Model ini dibangun menggunakan 300 pohon keputusan (n_estimators=300) dengan kedalaman maksimum 12 (max_depth=12) untuk mencegah pemasangan yang berlebihan. Parameter min_samples_split=4 dan

`min_samples_leaf=2` memastikan bahwa setiap simpul memiliki jumlah sampel yang cukup sebelum membelah atau menjadi daun. Jumlah fitur yang dipertimbangkan pada setiap pemisahan diatur dengan `max_features='sqrt'`, sedangkan `class_weight='seimbang'` digunakan untuk menangani ketidakseimbangan jumlah sampel di seluruh kelas. Pengaturan `random_state=42` memastikan hasil yang dapat direproduksi, dan `n_jobs=-1` mempercepat proses pelatihan. Konfigurasi ini menghasilkan model stabil yang mampu memberikan prediksi yang akurat untuk kelima kelas target.

E. Evaluasi Hasil Model

Setelah melatih model Random Forest, langkah selanjutnya adalah mengevaluasi kinerjanya. Evaluasi dilakukan dengan menghitung metrik seperti akurasi, presisi, ingatan, dan skor F1, serta menggunakan matriks kebingungan untuk memberikan gambaran komprehensif tentang hasil prediksi model dengan membandingkan label yang diprediksi dengan label aktual dari himpunan data pengujian. Pendekatan ini memungkinkan penilaian kinerja model yang jelas dan menyeluruh tanpa melibatkan pemodelan tambahan atau langkah penyetyelan hiperparameter.



Gambar 2 Evaluasi Model

Bagan menampilkan metrik evaluasi model Random Forest dalam bagan batang, termasuk Akurasi 96.13%, Presisi 96.29%, Recall 96.13%, dan Skor F-1 Score 96.17%. Dari grafik dapat dilihat bahwa semua metrik berada di kisaran 96%, menunjukkan bahwa model Random Forest memiliki kinerja yang stabil dan baik dalam memprediksi lima kelas target. Nilai presisi dan penarikan yang tinggi menunjukkan bahwa model mampu meminimalkan kesalahan prediksi, baik untuk kasus positif maupun negatif, menghasilkan Skor F1 yang tinggi, yang merupakan rata-rata harmonik dari presisi dan penarikan. Di sebaliknya, ada matriks kebingungan yang menunjukkan kemampuan model untuk membedakan antara data aktual dan prediksi, memungkinkan verifikasi akurasi deteksi untuk setiap kelas. Kesimpulannya, model Random Forest yang digunakan cukup akurat dan seimbang dalam menangani klasifikasi 5 kelas, yaitu rentan Tidak Miskin, Menengah, Hampir Miskin, Miskin, dan Sangat Miskin.

F. Analisis Interpretabilitas (SHAP) dan Rekomendasi

Analisis interpretabilitas menggunakan SHAP (SHapley Additive exPlanations) dilakukan untuk memahami kontribusi setiap fitur terhadap hasil klasifikasi tingkat kemiskinan oleh model Random Forest. Hasil analisis menunjukkan bahwa tiga fitur paling berpengaruh adalah Persentase Penduduk Miskin (P0), Pengeluaran per Kapita Disesuaikan, dan Rata-rata Lama Sekolah Penduduk Usia 15 Tahun ke Atas. Secara global, nilai P0 yang tinggi meningkatkan kemungkinan suatu wilayah dikategorikan sebagai “Miskin” atau “Sangat Miskin”, sedangkan pengeluaran per kapita dan lama sekolah yang lebih tinggi menurunkan risiko tersebut. Analisis lokal memperlihatkan bahwa pada wilayah tertentu, kombinasi rendahnya tingkat pendidikan dan tingginya P0 menjadi faktor dominan yang mendorong prediksi kemiskinan.

Rumus persamaan untuk mendapatkan nilai SHAP Indicator Feature Importance :

1. Untuk mendapatkan nilai rata – rata nilai absolut SHAP

$$M_{\{j,k\}} = \frac{1}{N} \sum_{i=1}^N shap[i, j, k] \quad (1)$$

Keterangan :

$M_{\{j,k\}}$ = Kontribusi rata-rata nilai dari fitur ke-j untuk kelas k
N = Jumlah Sampel
 $Shap_{[i,j,k]}$ = nilai SHAP untuk sampel i, fitur j, kelas k
 Σ = Nilai absolut digunakan untuk mengukur besar pengaruh, bukan arah pengaruh

2. Untuk mendapatkan rata-rata seluruh kelas menjadi 1 nilai per fitur

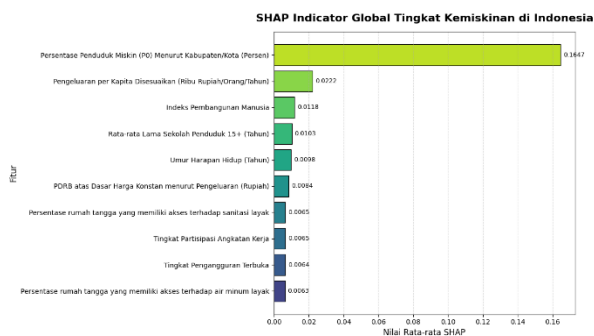
$$SHAP_Global_{\{j\}} = \frac{1}{N} \sum_{k=1}^c M_{j,k} \quad (2)$$

Keterangan :

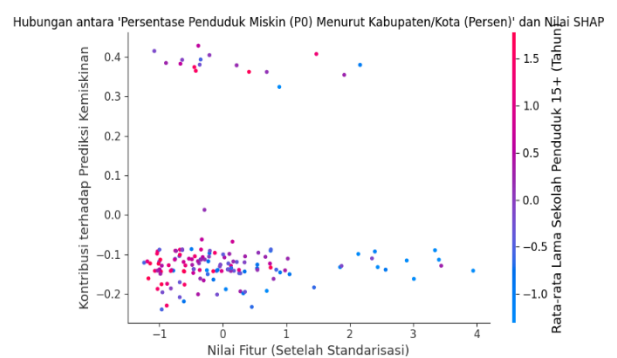
N = Jumlah sampel dalam data pelatihan
F = Jumlah fitur
C = Jumlah kelas (multi-class)
 $Shap_{[i,j,k]}$ = Nilai SHAP untuk sampel ke-i, fitur ke-j, kelas ke-k
 $M_{j,k}$ = Rata-rata absolut SHAP untuk fitur j pada kelas k
 $SHAP_Global_j$ = Nilai pentingnya fitur j secara global

Tabel 3. SHAP Indicator Feature Importance

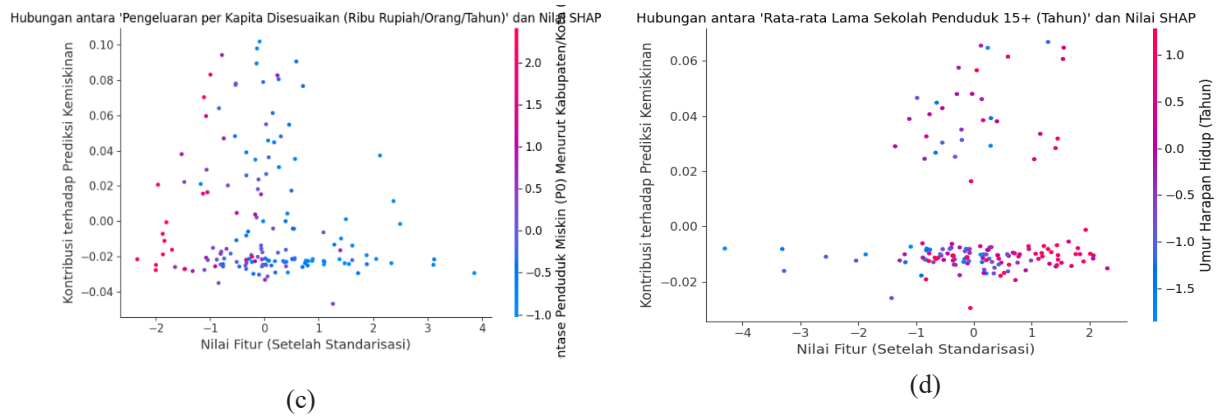
Fitur	SHAP Indicator
Persentase Penduduk Miskin (P0) Menurut Kabupaten/Kota (Persen)	0.164676
Rata-rata Lama Sekolah Penduduk 15+ (Tahun)	0.010254
Pengeluaran per Kapita Disesuaikan (Ribuan Rupiah/Orang/Tahun)	0.022190
Indeks Pembangunan Manusia	0.011776
Umur Harapan Hidup (Tahun)	0.009783
Persentase rumah tangga yang memiliki akses terhadap sanitasi layak	0.006488
Persentase rumah tangga yang memiliki akses terhadap air minum layak	0.006335
Tingkat Pengangguran Terbuka	0.006363
Tingkat Partisipasi Angkatan Kerja	0.006458
PDRB atas Dasar Harga Konstan menurut Pengeluaran (Rupiah)	0.008418



(a)



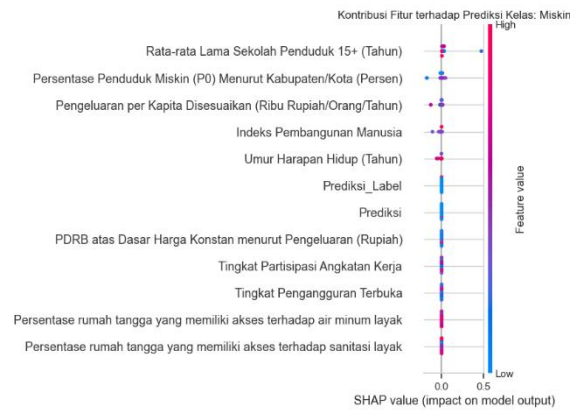
(b)



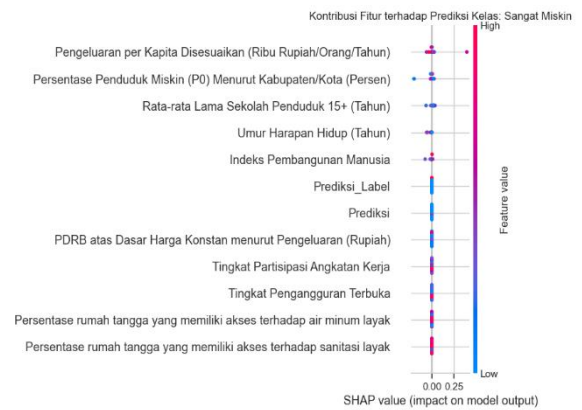
Gambar 3. Interpretabilitas SHAP Nilai Fitur

Analisis lokal menunjukkan bahwa fitur spesifik secara signifikan mempengaruhi prediksi wilayah tertentu terhadap kelas kemiskinan yang lebih tinggi. Berdasarkan temuan tersebut, intervensi yang direkomendasikan antara lain meningkatkan pendidikan, mengembangkan infrastruktur sanitasi dan air bersih, memberikan dukungan pendapatan, dan menerapkan program pengurangan pengangguran. Mengintegrasikan SHAP memungkinkan model ini tidak hanya akurat tetapi juga memberikan wawasan yang transparan dan dapat ditindaklanjuti bagi pembuat kebijakan untuk merancang intervensi berbasis data yang disesuaikan dengan kondisi sosial-ekonomi masing-masing wilayah.





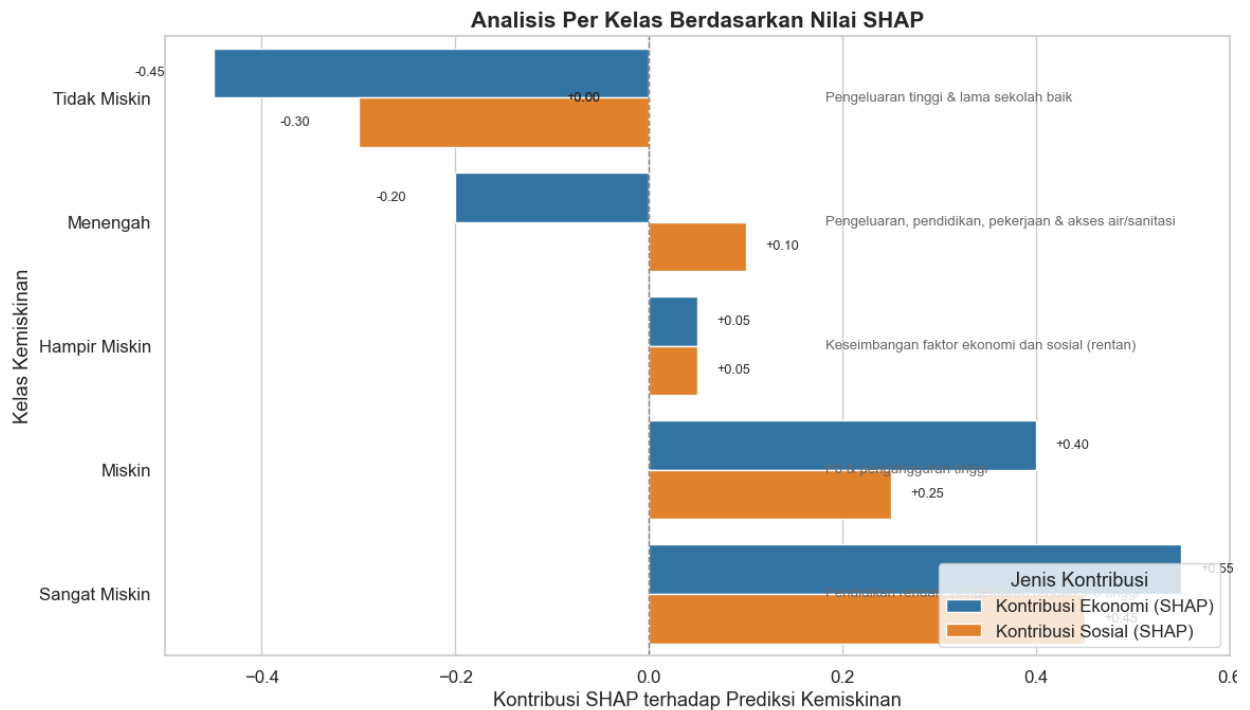
(e) Shape Kelas Miskin



(f) Shape Sangat Miskin

Gambar 4 Analisis Menggunakan Machine Learning Dengan Pendekatan SHAPE

Gambar 4 menampilkan hasil analisis SHAP pada model Random Forest dalam mengklasifikasikan tingkat kemiskinan di Indonesia. Secara umum, setiap kelas menunjukkan pola kontribusi fitur yang berbeda. Pada kelas Tidak Miskin, wilayah cenderung memiliki pengeluaran dan pendidikan tinggi, dengan nilai SHAP negatif yang menandakan penurunan risiko kemiskinan. Kelas Menengah masih dipengaruhi oleh faktor indeks pembangunan manusia dan umur harapan hidup penduduk, namun mulai terlihat peran indikator pekerjaan serta akses terhadap fasilitas dasar seperti air bersih dan sanitasi. Pada kelas Hampir Miskin, terdapat keseimbangan antara kontribusi positif dan negatif dari faktor ekonomi dan sosial, menunjukkan kondisi ekonomi yang rentan dan mudah berubah akibat faktor eksternal. Sementara itu, kelas Miskin didominasi oleh tingginya persentase penduduk miskin (P0) dan tingkat pengangguran, sedangkan kelas Sangat Miskin dipengaruhi oleh kombinasi rendahnya pendidikan, rendahnya pengeluaran per kapita, serta tingginya nilai P0.



Gambar 5. Visualisasi Kelas

Pada level lokal, analisis SHAP menunjukkan adanya variasi antarwilayah, misalnya daerah dengan P0 tinggi dan pendidikan rendah tetap dikategorikan miskin meskipun memiliki pengeluaran yang relatif sedang. Secara keseluruhan, penggunaan SHAP membuat model Random Forest yang kompleks menjadi lebih transparan dan mudah diinterpretasikan, sehingga hasil klasifikasi tidak hanya akurat, tetapi juga dapat memberikan wawasan yang bermakna bagi perumusan kebijakan pengentasan kemiskinan.

Berdasarkan hasil interpretabilitas menggunakan SHAP, beberapa rekomendasi kebijakan dapat dirumuskan untuk menangulangi kemiskinan secara lebih terarah dan berbasis data. Pertama, diperlukan peningkatan akses dan kualitas

pendidikan, mengingat lama sekolah terbukti berperan signifikan dalam menurunkan risiko kemiskinan. Investasi berkelanjutan pada sarana pendidikan dasar dan menengah menjadi langkah strategis untuk memperkuat modal manusia. Kedua, pemberdayaan ekonomi dan peningkatan pendapatan perlu dilakukan melalui program bantuan usaha mikro, pelatihan keterampilan kerja, serta pengembangan ekonomi lokal yang dapat meningkatkan pengeluaran per kapita rumah tangga. Ketiga, peningkatan infrastruktur dasar seperti penyediaan sanitasi layak, air bersih, dan akses transportasi menjadi faktor penting yang berkorelasi positif dengan kesejahteraan masyarakat. Terakhir, pendekatan kebijakan berbasis data perlu diterapkan dengan memanfaatkan hasil analisis SHAP untuk mengidentifikasi wilayah-wilayah dengan faktor risiko dominan seperti rendahnya tingkat pendidikan atau pengeluaran, sehingga intervensi dapat dilakukan secara lebih tepat sasaran dan efektif dalam menurunkan tingkat kemiskinan.

Kesimpulan

Penelitian ini berhasil mengembangkan model klasifikasi tingkat kemiskinan di Indonesia dengan memanfaatkan algoritma Random Forest yang dikombinasikan dengan metode Explainable AI (SHAP) untuk memberikan interpretasi yang transparan terhadap hasil prediksi. Model yang dibangun menunjukkan kinerja yang sangat baik dengan akurasi sebesar 96,13%, presisi 96,29%, recall 96,13%, dan F1-score 96,17%. Hasil ini membuktikan bahwa Random Forest mampu menangani kompleksitas data sosial ekonomi dengan stabilitas dan akurasi tinggi. Analisis interpretabilitas menggunakan SHAP mengidentifikasi tiga fitur paling berpengaruh terhadap klasifikasi kemiskinan, yaitu persentase penduduk miskin (P0), pengeluaran per kapita, dan rata-rata lama sekolah. Secara global, peningkatan pendidikan dan pengeluaran per kapita terbukti menurunkan tingkat kemiskinan, sementara tingginya P0 dan pengangguran menjadi indikator dominan wilayah miskin dan sangat miskin. Integrasi SHAP dalam penelitian ini memberikan nilai tambah berupa kemampuan untuk memahami kontribusi masing-masing variabel terhadap hasil model, sehingga mendukung perumusan kebijakan berbasis data yang lebih efektif dan tepat sasaran. Hasil penelitian ini diharapkan dapat menjadi acuan bagi pemerintah dan lembaga terkait dalam merancang program pengentasan kemiskinan yang berfokus pada peningkatan kualitas pendidikan, pemberdayaan ekonomi, serta pembangunan infrastruktur dasar.

Untuk penelitian selanjutnya, disarankan agar dilakukan pengembangan model dengan metode ensemble yang lebih kompleks seperti XGBoost, CatBoost, atau LightGBM, serta integrasi data spasial dan temporal untuk meningkatkan akurasi prediksi. Selain itu, penerapan pendekatan Explainable AI multimetode dapat memperkaya interpretasi dan memperkuat keandalan hasil dalam konteks kebijakan sosial ekonomi di Indonesia.

Daftar Pustaka

- [1] Badan Pusat Statistik, “Statistik Kemiskinan Indonesia 2023,” 2023.
- [2] World Bank, “Indonesia Economic Prospects: Boosting the Recovery,” 2022.
- [3] L. Khikmah dan M. Sofro, “Multidimensional Poverty Measurement and its Implications in Policy,” *J. Dev. Stud.*, vol. 61, no. 2, hal. 210–225, 2025, doi: 10.1080/00220388.2025.00123.
- [4] B. R. Lamichhane, “Exploring machine learning trends in poverty mapping,” *J. Spat. Sci.*, vol. 66, no. 1, hal. 1–20, 2025, doi: 10.1080/14498596.2025.2000069.
- [5] R. Putra, “Socioeconomic Indicators and Poverty Mapping Using Random Forest,” *Int. J. Data Sci.*, vol. 9, no. 2, hal. 88–101, 2024, doi: 10.1016/j.ijdatasci.2024.05.008.
- [6] J. Jeffry, N. Nurdyansa, S. Usman, D. Sasmita, dan M. N. Arafah, “Transformasi Digital Dalam Industri Kerajinan Rumahan Melalui Penggunaan Teknologi Machine Learning,” *JMM (Jurnal Masy. Mandiri)*, vol. 7, no. 6, hal. 5840–5849, 2023.
- [7] D. Lasfeto, H. Lestari, dan A. Pratama, “Integrating Regional Economic Indicators for Poverty Assessment in Indonesia,” *Indones. J. Econ. Dev.*, vol. 21, no. 1, hal. 50–66, 2024.
- [8] R. Khalisha, “Random Forest for Multivariate Socioeconomic Data Analysis,” *J. Intell. Syst.*, vol. 33, no. 7, hal. 870–882, 2024.
- [9] D. Ramayanti, A. Raharjo, dan S. Yuliani, “Classification of Poverty Levels Using Random Forest Approach,” *J. Appl. Econ. Anal.*, vol. 45, no. 4, hal. 321–338, 2023.
- [10] T. Santos dan P. Almeida, “Poverty Mapping and Machine Learning: A Review of Methods and Challenges,” *Soc. Indic. Res.*, vol. 150, no. 3, hal. 769–790, 2020.
- [11] M. Chulafak, F. Karim, dan D. Wibowo, “Spatial Modeling of Poverty Using Random Forest in Indonesia,” *Geoj. Dev. Stud.*, vol. 18, no. 2, hal. 201–218, 2024.
- [12] R. Gulecha, A. Singh, dan V. Kumar, “Using Machine Learning for Socioeconomic Poverty Prediction,” *Data Policy*, vol. 6, no. 1, hal. 1–12, 2024, doi: 10.1017/dap.2024.12.

- [13] S. Putri, N. Hidayat, dan R. Fadli, "Feature Importance in Random Forest for Poverty Prediction," *Procedia Comput. Sci.*, vol. 199, hal. 1215–1222, 2022, doi: 10.1016/j.procs.2022.01.150.
- [14] G. Watmough, C. Huggins, dan R. Osei, "Explainable AI for Poverty Analysis Using SHAP Values," *Artif. Intell. Dev.*, vol. 8, no. 2, hal. 55–72, 2025.
- [15] M. El-Geneedy, R. Ahmed, dan F. Salim, "Transparency in Poverty Classification through Explainable AI," *Comput. Soc.*, vol. 45, no. 1, hal. 33–49, 2025.
- [16] J. Hall, L. Thomas, dan T. Nguyen, "AI and Sustainable Development: Policy Applications and Future Directions," *Glob. Policy J.*, vol. 14, no. 3, hal. 201–217, 2023.