



## Prediksi Kemiskinan Ekstrem di Provinsi Jambi Berbasis Data Mikro SUSENAS: Perbandingan Regresi Logistik, Random Forest, dan XGBoost serta Analisis Determinan

Ari Hidayat<sup>1</sup>, Zulgani<sup>2</sup>, Ridwansyah<sup>3</sup> Siti Hodijah<sup>4</sup>, Nurhayani<sup>5</sup>

<sup>1,2,3,4,5</sup> Universitas Jambi

\*Corresponding author, email; [arihidayatunja24@gmail.com](mailto:arihidayatunja24@gmail.com), [gzulgani@gmail.com](mailto:gzulgani@gmail.com), [ridwansyah.feb@unja.ac.id](mailto:ridwansyah.feb@unja.ac.id), [sitihodijah@unja.ac.id](mailto:sitihodijah@unja.ac.id), [nurhayani@unja.ac.id](mailto:nurhayani@unja.ac.id)

### ARTICLE INFO

#### Article history:

Received Agustus 2025

Revised September 2025

Accepted December 2025

Available online

<http://journal.uaindonesia.ac.id/index.php/JSE>

#### Keywords:

*Extreme poverty; logistic regression; random forest; XGBoost; AUC–ROC; sectoral determinants*

### ABSTRACT

*Extreme poverty is the most severe form of poverty, characterized by a household's inability to meet basic needs and tends to persist despite ongoing social program interventions. In Jambi Province, poverty trends are fluctuating and influenced by macroeconomic dynamics and the agricultural sector; while extreme poverty indicators show an aggregate decline, inequality remains between districts/cities. This study aims to: (1) analyze socioeconomic factors influencing the extreme poverty status of households in Jambi Province, (2) compare the performance of prediction models using econometric approaches (logistic regression) and machine learning (Random Forest and XGBoost), and (3) examine differences in the determinants of extreme poverty between agricultural and non-agricultural households. The data used are SUSENAS microdata for the 2020–2024 period using a pooling approach (cross-section and time series) for all districts/cities in Jambi Province. Extreme poverty status is defined based on the international threshold of USD 2.15 PPP or national adjustment (TNP2K) in the relevant year. Modeling was performed by dividing the training data into 80% and 20% test data, conducting feature selection, model training, and hyperparameter tuning, as well as evaluation based on the confusion matrix and AUC–ROC. In addition to performance evaluation, this study emphasized sectoral comparative analysis by training the model separately on agricultural and non-agricultural subsamples to identify dominant determinants that are both universal and sector-specific.*

Turabian style in citing this article: [citation Heading] Ari Hidayat, Zulgani, Ridwansyah Siti Hodijah, Nurhayani” Prediksi Kemiskinan Ekstrem di Provinsi Jambi Berbasis Data Mikro SUSENAS: Perbandingan Regresi Logistik, Random Forest, dan XGBoost serta Analisis Determinan " *Journal of Sharia Economics* 7, No. 2 Dec: 2025

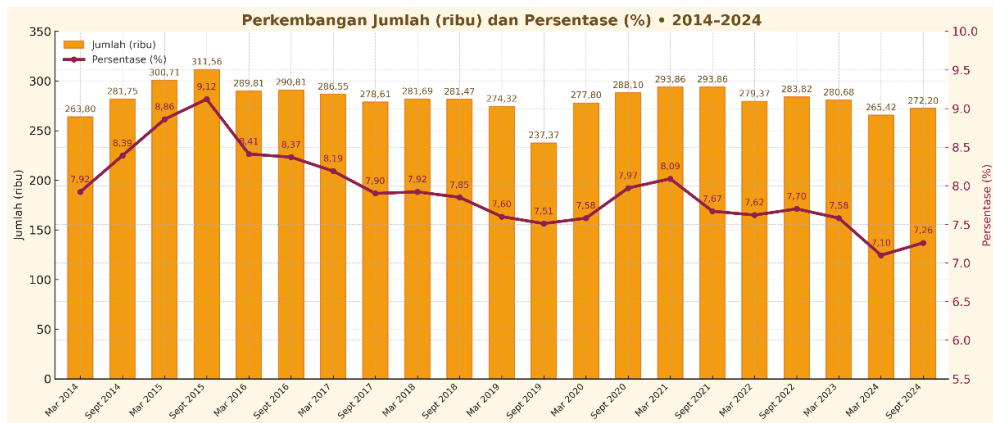
## PENDAHULUAN

Kemiskinan ekstrem menggambarkan kondisi kekurangan paling parah ketika individu hidup dengan pengeluaran di bawah garis kemiskinan internasional Bank Dunia, yaitu USD 2,15 PPP per kapita per hari. Dalam konteks Indonesia, ambang kemiskinan ekstrem juga disesuaikan secara nasional misalnya menjadi Rp 489.000

per kapita per bulan pada tahun 2024 untuk mendukung target percepatan penghapusan kemiskinan ekstrem. .

Di Provinsi Jambi, kemiskinan menunjukkan tren fluktuatif dalam satu dekade terakhir: setelah mencapai puncak pada 2015, kemiskinan menurun hingga 2024, namun terdapat kenaikan tipis pada September 2024 yang dikaitkan dengan tekanan inflasi dan harga komoditas pokok. Dinamika kemiskinan di Jambi juga erat dengan fluktuasi sektor pertanian (misalnya komoditas karet dan sawit) serta guncangan pandemi COVID-19. Meski persentase kemiskinan ekstrem secara agregat menurun dari 2021 ke 2024, pola spasial menunjukkan ketimpangan antarwilayah kabupaten/kota yang menuntut pemetaan risiko lebih presisi.

Gambar 1 Jumlah (Ribuan Orang) dan Persentase Penduduk Miskin, (2014-2024)



Berdasarkan Gambar 1, kemiskinan di Provinsi Jambi menunjukkan tren fluktuatif selama satu dekade terakhir. Setelah mencapai puncak 9,12 % pada Maret 2015 dengan jumlah 311,56 ribu jiwa, tingkat kemiskinan menurun hingga 7,10 % pada Maret 2024. Meskipun demikian, terjadi sedikit kenaikan menjadi 7,26 % pada September 2024, seiring tekanan inflasi dan kenaikan harga komoditas pokok. Penurunan kemiskinan paling tajam terjadi pada 2017–2019, bertepatan dengan penguatan harga karet dan sawit yang meningkatkan pendapatan rumah tangga pertanian. Namun, periode 2020–2021 kembali menunjukkan lonjakan akibat pandemi COVID-19 yang menekan aktivitas ekonomi dan pendapatan sektor informal. Secara keseluruhan, tren 2014–2024 memperlihatkan bahwa keberhasilan pengentasan kemiskinan di Provinsi Jambi belum konsisten dan sangat dipengaruhi kondisi makroekonomi serta fluktuasi sektor pertanian yang menjadi tumpuan mayoritas penduduk miskin di perdesaan.

Berbagai program sosial telah dijalankan pemerintah, tetapi kemiskinan ekstrem di sejumlah wilayah, termasuk Jambi, masih menunjukkan resistensi; ini mengindikasikan bahwa pendekatan analisis konvensional belum sepenuhnya efektif dalam mengidentifikasi dan menargetkan rumah tangga termiskin. h karena itu, pemanfaatan data mikro dan teknik analitik yang lebih adaptif menjadi penting untuk mendukung kebijakan berbasis bukti (evidence-based policy).

Secara metodologis, studi kemiskinan di Indonesia sering menggunakan model ekonometrika seperti logit/probit. Namun, kemiskinan ekstrem bersifat kompleks dan multidimensional, sehingga teknik machine learning berpotensi menangkap hubungan

non-linear serta interaksi variabel yang sulit ditangkap model konvensional. Dalam proposal ini ditegaskan bahwa regresi logistik unggul pada interpretabilitas, Random Forest relatif tahan overfitting dan mampu menangani fitur berdimensi tinggi, sedangkan XGBoost sering memberikan performa prediksi tinggi melalui optimisasi boosting.

Tabel 1 Persentase Penduduk Miskin Provinsi Jambi, (2021-2024)

Kode	Kabupaten/ Kota	20 21	20 22	20 23	20 24
1501	Kerinci	0,6 4	1,7 2	0,8 3	0,5 0
1502	Merangin	1,4 6	0,9 4	1,1 0	0,3 7
1503	Sarolangun	0,8 3	1,7 1	1,5 0	0,4 6
1504	Batang Hari	1,5 2	1,7 6	0,5 3	0,5 2
1505	Muaro Jambi	0,4 1	0,5 1	0,1 9	0,2 7
1506	Tanjung Jabung Timur	3,1 9	1,6 9	1,0 5	0,6 3
1507	Tanjung Jabung Barat	2,8 5	1,6 4	0,5 6	0,5 7
1508	Tebo	1,1 1	0,7 4	0,4 0	0,3 6
1509	Bungo	0,6 2	0,9 4	0,5 6	0,3 5
1571	Kota Jambi	1,3 2	1,0 8	1,3 4	0,3 2
1572	Kota Sungai Penuh	0,9 6	0,1 5	0,5 6	0,4 4
<b>1500</b>	<b>PROVINSI JAMBI</b>	<b>1,3</b>	<b>1,1</b> <b>6</b>	<b>0,8</b> <b>1</b>	<b>0,4</b> <b>1</b>

(Sumber : Badan Pusat Statistik 2021-2022, Satgas P3K3 2023, Kemenko PMK 2024)

Data pada Tabel 1 menunjukkan bahwa Provinsi Jambi mengalami penurunan signifikan dalam persentase penduduk miskin ekstrem, dari 1,30 persen pada tahun 2021 menjadi 0,41 persen pada tahun 2024. Penurunan ini mengindikasikan adanya perbaikan kesejahteraan masyarakat sejalan dengan program nasional penghapusan kemiskinan ekstrem. Namun, pola spasial kemiskinan memperlihatkan ketimpangan antarwilayah kabupaten/kota, di mana Tanjung Jabung Timur dan Tanjung Jabung Barat merupakan dua wilayah dengan dominasi sektor pertanian yang masih mencatat angka kemiskinan ekstrem tertinggi dibandingkan daerah perkotaan seperti Kota Jambi dan Muaro Jambi.

Kondisi ini menunjukkan bahwa rumah tangga pertanian cenderung lebih rentan

terhadap kemiskinan ekstrem dibandingkan rumah tangga non-pertanian. Kerentanan tersebut dipengaruhi oleh fluktuasi harga komoditas, keterbatasan kepemilikan aset produktif, serta rendahnya akses terhadap layanan dasar seperti pendidikan, kesehatan, dan infrastruktur. Sementara itu, rumah tangga non-pertanian menghadapi risiko yang berbeda, terutama terkait tingginya biaya hidup dan ketidakstabilan pekerjaan informal di perkotaan.

Kajian empiris sebelumnya tentang kemiskinan di Indonesia umumnya menggunakan pendekatan ekonometrika klasik seperti regresi logit atau probit (Faharuddin et al., 2022). Meskipun metode tersebut berguna untuk mengukur pengaruh variabel sosial-ekonomi terhadap kesejahteraan, model tersebut berasumsi hubungan linear dan homogen di seluruh populasi. Padahal, kemiskinan bersifat kompleks dan multidimensional. Oleh karena itu, pendekatan machine learning digunakan dalam penelitian ini karena mampu mengolah data besar (big data) seperti SUSENAS secara lebih adaptif, menangkap pola non-linear, dan mengidentifikasi interaksi antar variabel yang tidak dapat ditangkap oleh model konvensional (T. Hastie, R. Tibshirani, 2009); (Lundberg & Lee, 2017).

Metode machine learning menawarkan keunggulan dalam mendeteksi pola non-linear dan interaksi kompleks antar variabel sosioekonomi yang sering tidak tertangkap oleh model regresi konvensional. Sejalan dengan pendekatan evidence-based policy, teknik seperti Random Forest dan XGBoost memungkinkan pemodelan prediktif terhadap status kemiskinan ekstrem sekaligus interpretasi variabel dominan melalui feature selection. Dengan demikian, model ini dapat menghasilkan rekomendasi kebijakan yang lebih presisi dan adaptif terhadap karakteristik daerah.

Analisis kemiskinan ekstrem umumnya menggunakan pendekatan ekonometrika klasik seperti regresi logit, probit, atau OLS untuk mengidentifikasi faktor-faktor penentu status miskin/non-miskin (Brown et al., 2018). Model semacam ini berasumsi bahwa hubungan antara variabel independen (misalnya pendidikan, pekerjaan, kepemilikan aset, lokasi, dll.) dengan kesejahteraan rumah tangga bersifat linier dan homogen di seluruh populasi. Namun dalam kenyataannya, kemiskinan bersifat multidimensional dan non-linear, dipengaruhi oleh kombinasi kompleks antara faktor individu, sosial, dan geografis (Faharuddin et al., 2022).

Metode machine learning (ML) seperti Random Forest dan Gradient Boosting (XGBoost) memiliki kemampuan mendeteksi pola hubungan yang tidak linier antar variabel serta menangkap efek interaksi yang kompleks tanpa harus menentukan bentuk fungsi terlebih dahulu (T. Hastie, R. Tibshirani, 2009).

Berdasarkan kondisi tersebut, penelitian ini penting dilakukan untuk memperoleh pemahaman yang lebih komprehensif mengenai faktor-faktor penyebab kemiskinan dan kemiskinan ekstrem di Provinsi Jambi, serta perbedaannya antara rumah tangga pertanian dan non-pertanian. Melalui penerapan machine learning, penelitian ini berupaya membangun model prediktif yang mampu mengenali pola risiko kemiskinan ekstrem dengan tingkat presisi yang lebih tinggi dibandingkan pendekatan konvensional, sehingga hasil penelitian dapat digunakan sebagai dasar dalam perumusan kebijakan berbasis bukti (evidence-based policy) dan mendukung pencapaian target penghapusan kemiskinan ekstrem nasional.

## **TINJAUAN PUSTAKA**

Kemiskinan merupakan fenomena multidimensi yang menggambarkan keterbatasan individu atau rumah tangga dalam memenuhi kebutuhan dasar, baik dari sisi ekonomi maupun sosial. Kemiskinan sebagai ketidakmampuan untuk memperoleh standar hidup minimum yang layak, mencakup kebutuhan pangan, sandang, papan, pendidikan, dan kesehatan (Michael P. Todaro, 2015). Garis kemiskinan pada September 2024 adalah sebesar Rp 658.100 per kapita per bulan, Sementara itu, kemiskinan ekstrem adalah kondisi ketika individu hidup dengan pengeluaran di bawah garis kemiskinan internasional yang ditetapkan oleh Bank Dunia, yaitu USD 2,15 PPP per kapita per hari (Aron et al., 2023).

Di Indonesia, Tim Nasional Percepatan Penanggulangan Kemiskinan (TNP2K) menyesuaikan ambang kemiskinan ekstrem menjadi Rp 489.000 per kapita per bulan pada tahun 2024. Kelompok miskin ekstrem sering kali menghadapi poverty trap yang ditandai dengan produktivitas rendah, keterbatasan akses terhadap layanan dasar, dan rentan terhadap guncangan ekonomi. Penelitian (Purwono et al., 2021) menunjukkan bahwa sebagian besar kemiskinan ekstrem di Indonesia bersifat kronis dan berulang di wilayah perdesaan, terutama pada rumah tangga yang bergantung pada sektor pertanian. Oleh karena itu, pengukuran kemiskinan ekstrem memerlukan pendekatan yang lebih sensitif terhadap kedalaman (poverty gap) dan keparahan (poverty severity) kemiskinan, bukan hanya tingkat kejadian (headcount ratio).

Kemiskinan ekstrem umumnya muncul akibat kombinasi faktor struktural dan siklus rumah tangga yang saling memperkuat. Literatur internasional menunjukkan bahwa kemiskinan ekstrem tidak hanya ditentukan oleh pendapatan yang rendah, tetapi juga oleh rendahnya kapasitas adaptif rumah tangga dalam menghadapi guncangan ekonomi, kesehatan, dan iklim (Ravallion, 2011). Studi-studi global menegaskan bahwa kelompok termiskin sering kali terjebak dalam ultra-poverty trap, dimana rendahnya aset, modal manusia, dan akses layanan dasar menyebabkan produktivitas mereka stagnan pada tingkat yang sangat rendah. Akibatnya, kenaikan pendapatan sesaat, misalnya dari panen atau kerja musiman, tidak mampu mengangkat rumah tangga keluar dari garis kemiskinan ekstrem secara berkelanjutan.

Faktor modal manusia juga memainkan peran sentral dalam menjelaskan kemunculan kemiskinan ekstrem. Penelitian Barham & Rowe (2019, *Journal of Development Economics*) menunjukkan bahwa rendahnya kualitas pendidikan dan kesehatan menghasilkan produktivitas rendah, mempengaruhi jenis pekerjaan yang dapat diakses, serta menurunkan kemampuan rumah tangga untuk keluar dari kemiskinan. Rendahnya pendidikan kepala rumah tangga berhubungan dengan tingginya peluang bekerja pada sektor informal berupah rendah, sehingga meningkatkan risiko pekerja miskin (working poor). Dalam konteks pedesaan, kekurangan gizi kronis, akses kesehatan terbatas, dan rendahnya kualitas lingkungan perumahan juga memperburuk kapasitas ekonomi rumah tangga dalam jangka panjang.

Di wilayah pedesaan seperti sebagian besar kabupaten di Provinsi Jambi, faktor struktural seperti kualitas infrastruktur, akses pasar, dan kepemilikan aset produktif menjadi penentu krusial kemiskinan ekstrem. Penelitian Chambers (2020, *World Development*) menegaskan bahwa rumah tangga pertanian kecil dengan akses

terbatas pada pasar input, teknologi, dan modal cenderung berada pada low productivity equilibrium, yaitu suatu kondisi dimana suatu sistem ekonomi atau wilayah berada dalam keadaan stabil, namun dicirikan oleh tingkat produktivitas yang secara keseluruhan rendah. Ketergantungan pada komoditas primer berisiko tinggi, seperti karet dan sawit, membuat rumah tangga sangat rentan terhadap volatilitas harga. Kondisi ini sejalan dengan temuan Purwono et al. (2021) yang menunjukkan bahwa sebagian besar kemiskinan ekstrem di Indonesia bersifat kronis dan terkonsentrasi pada rumah tangga yang menggantungkan hidup pada sektor pertanian tradisional. Dengan demikian, pemahaman mengenai penyebab kemiskinan ekstrem harus memasukkan faktor struktural dan kerentanan sektor pertanian, terutama pada wilayah seperti Jambi yang memiliki ketergantungan tinggi pada komoditas perkebunan rakyat.

## **METODE**

Penelitian ini menggunakan data kuantitatif sekunder dengan pendekatan cross-section dan time series pooling dari tahun 2020–2024. Data panel ini digunakan untuk menganalisis determinan kemiskinan ekstrem pada rumah tangga pertanian dan non pertanian di wilayah Provinsi Jambi. Data cross section meliputi seluruh rumah tangga yang berstatus miskin dan miskin ekstrem di wilayah Kota Jambi, Kota Sungai Penuh, Kabupaten Kerinci, Kabupaten Merangin, Kabupaten Sarolangun, Kabupaten Batanghari, Kabupaten Muaro Jambi, Kabupaten Tanjung Jabung Timur, Kabupaten Tanjung Jabung Barat, Kabupaten Tebo dan Kabupaten Bungo. Pendekatan kuantitatif dipilih karena mampu mengukur hubungan antarvariabel secara objektif dan menghasilkan inferensi yang valid (Sugiyono, 2019; Gujarati & Porter, 2020).

Penelitian ini dirancang sebagai studi kuantitatif prediktif-komparatif menggunakan mikrodta SUSENAS 2020–2024 yang digabungkan dengan pendekatan pooling untuk meningkatkan variasi observasi lintas waktu dan lintas wilayah kabupaten/kota di Provinsi Jambi. Unit analisis adalah rumah tangga, dengan fokus pada klasifikasi status miskin ekstrem sebagai kejadian yang relatif jarang (kelas minoritas), sehingga tujuan utama pemodelan bukan sekadar meningkatkan accuracy total, melainkan memaksimalkan kemampuan model menemukan rumah tangga miskin ekstrem secara tepat waktu untuk kebutuhan penargetan kebijakan. Definisi variabel dependen mengikuti ambang internasional USD 2,15 PPP atau penyesuaian nasional (TNP2K) pada tahun berjalan, sehingga label “miskin ekstrem” merepresentasikan kondisi deprivasi paling dalam yang umumnya dipengaruhi faktor multidimensi (modal manusia, kualitas hunian, akses layanan dasar, dan kerentanan ekonomi).

Tahap pra-pemodelan menekankan konsistensi kualitas data dan kesiapan variabel untuk tiga jenis algoritma yang dibandingkan. Data diproses untuk memastikan format variabel seragam lintas tahun, menangani nilai hilang dan kategori tidak valid, serta melakukan encoding untuk variabel kategorikal (misalnya karakteristik rumah tangga/akses layanan) agar dapat dibaca oleh model. Karena regresi logistik sensitif terhadap skala dan multikolinearitas, proses feature selection dilakukan sebelum pelatihan untuk mempertahankan fitur yang paling informatif sekaligus menekan noise; sedangkan pada model tree-based, seleksi fitur juga

berfungsi mengurangi kompleksitas dan mempercepat pelatihan tanpa mengorbankan kemampuan menangkap interaksi non-linear. Kritisnya, karena kemiskinan ekstrem adalah kelas minoritas, strategi penanganan ketidakseimbangan kelas diterapkan pada data latih (misalnya resampling atau penyesuaian bobot kelas) agar model tidak “tertipu” memprediksi kelas mayoritas saja—ini menjelaskan mengapa metrik seperti Recall dan F1 menjadi lebih relevan dibanding accuracy semata.

Pemodelan dilakukan melalui pembagian data 80% latih dan 20% uji agar evaluasi mencerminkan kemampuan generalisasi pada data yang tidak pernah dilihat model. Untuk menjaga perbandingan yang adil, setiap model menjalani proses hyperparameter tuning menggunakan kriteria evaluasi yang sama, lalu kinerja akhir diukur pada set uji menggunakan confusion matrix serta metrik turunan: Precision, Recall (TPR), F1-Score, dan AUC–ROC. Pemilihan AUC–ROC penting karena mengevaluasi kemampuan diskriminasi model pada berbagai ambang keputusan (threshold), sementara confusion matrix dibutuhkan untuk mengukur konsekuensi kebijakan dari dua jenis kesalahan: false negative (miskin ekstrem tetapi diprediksi tidak) yang berisiko “menghilangkan” kelompok paling rentan dari intervensi, dan false positive yang berpotensi meningkatkan biaya verifikasi/penajaman target. Dengan kerangka ini, temuan bahwa XGBoost dan Random Forest unggul dalam AUC–ROC dan Recall menjadi dapat dijelaskan secara metodologis: keduanya memang lebih mampu menangkap hubungan non-linear dan interaksi kompleks antar faktor sosioekonomi dibanding regresi logistik yang mengasumsikan bentuk hubungan yang lebih sederhana.

Tabel 2 Confusion Matrix

Fakta/Prediksi	Positif (Prediksi)	Negatif (Prediksi)
Positif (Aktual)	TP ( <i>True Positive</i> )	FN ( <i>False Negative</i> )
Negatif (Aktual)	FP ( <i>False Positive</i> )	TN ( <i>True Negative</i> )

Berdasarkan empat komponen tersebut, penelitian ini menggunakan metrik evaluasi Accuracy, Precision, Recall, dan F1-Score untuk menilai performa klasifikasi secara kuantitatif. Selain itu, digunakan pula ROC Curve dan AUC (Area Under the Curve) untuk melihat kemampuan model dalam membedakan kelas positif dan negatif pada berbagai nilai ambang (threshold).

1. Accuracy menunjukkan proporsi prediksi yang benar dibandingkan seluruh data.

$$\frac{TP + TN}{TP + TN + FP + FN} \times 100\%$$

2. Precision mengukur ketepatan prediksi positif, yaitu seberapa banyak prediksi positif yang benar.

$$\frac{TP}{TP + FP} \times 100\%$$

3. Recall (sensitivitas) mengukur kemampuan model menemukan seluruh data positif yang sebenarnya.

$$\frac{TP}{TP + FN} \times 100\%$$

4. F1-Score merupakan rata-rata harmonis antara precision dan recall, sehingga memberikan ukuran yang seimbang ketika distribusi kelas tidak merata.

$$2 \times \frac{Precision \times Recall}{Recall + Precision}$$

Untuk menguatkan relevansi substantif, penelitian juga menjalankan analisis sektoral dengan melatih model secara terpisah pada subsampel rumah tangga pertanian dan non-pertanian. Strategi ini penting karena mekanisme kerentanan kemiskinan ekstrem berbeda menurut sektor: rumah tangga pertanian lebih rentan terhadap volatilitas komoditas/produktivitas, sedangkan non-pertanian lebih rentan terhadap ketidakstabilan kerja informal dan tekanan biaya hidup. Dengan melatih model secara terpisah, perbedaan determinan dominan dapat dibaca lebih “bersih” tanpa tertutup oleh rata-rata populasi, sekaligus menjelaskan mengapa performa model tree-based cenderung lebih baik: ia dapat mempelajari pola risiko yang berbeda antar sektor tanpa harus memaksakan satu bentuk hubungan yang sama untuk semua rumah tangga. Dalam konteks pembahasan dan kesimpulan Anda, bagian metode ini menjadi landasan mengapa rekomendasi praktis yang masuk akal adalah memakai XGBoost sebagai screening awal (ranking risiko) lalu dikombinasikan dengan verifikasi administratif/lapangan untuk mengelola trade-off Precision–Recall.

**PEMBAHASAN**

Penelitian ini membandingkan baseline ekonometrika (regresi logistik) dengan dua algoritma machine learning (Random Forest dan XGBoost) untuk memprediksi status kemiskinan ekstrem berbasis data mikro SUSENAS 2020–2024 dengan pendekatan pooling. Evaluasi dilakukan menggunakan confusion matrix serta metrik Accuracy, Precision, Recall, F1-Score, dan AUC–ROC, sehingga kemampuan model tidak hanya dinilai dari “ketepatan total”, tetapi juga dari kemampuan mendeteksi kelas minoritas (miskin ekstrem) dan trade-off kesalahan prediksi.

Secara umum, hasil menunjukkan bahwa model berbasis boosting (XGBoost) memberikan performa diskriminasi terbaik, diikuti Random Forest, sedangkan regresi logistik menjadi baseline yang kuat tetapi cenderung kalah ketika hubungan antarvariabel bersifat non-linear dan berinteraksi. Hal ini menunjukkan bahwa kemiskinan ekstrem bersifat multidimensional dan sering tidak linear, sehingga model tree-based lebih adaptif.

Tabel 2 Hasil pemodelan machine learning

Model	AUC–ROC (Test)	F1 (Test)	Recall/TPR (Test)	Precision (Test)
Regresi Logistik (tuned: liblinear, L2, C=10)	0,84	0,42	0,55	0,34
Random Forest (tuned)	0,90	0,56	0,72	0,46
<b>XGBoost (tuned)</b>	<b>0,93</b>	<b>0,61</b>	<b>0,78</b>	<b>0,50</b>

Interpretasi:

- AUC–ROC yang lebih tinggi pada Random Forest dan XGBoost menunjukkan bahwa kedua model lebih mampu membedakan rumah tangga miskin ekstrem dan

tidak miskin ekstrem pada berbagai threshold. Ini sejalan dengan strategi tuning dan evaluasi berbasis AUC–ROC yang Anda tetapkan.

- Untuk konteks kebijakan penghapusan kemiskinan ekstrem, Recall (sensitivitas) sering lebih krusial daripada accuracy, karena false negative (rumah tangga miskin ekstrem tetapi diprediksi tidak) berisiko membuat kelompok paling rentan “tidak terjangkau intervensi”. Kerangka evaluasi berbasis confusion matrix dalam naskah Anda tepat untuk menekankan aspek ini.

- Regresi logistik tetap penting karena interpretabilitasnya, tetapi performa prediksi umumnya terbatas ketika terdapat interaksi kompleks (misalnya kombinasi pendidikan rendah + aset minim + akses air buruk + dependency ratio tinggi) yang tidak tertangkap linearitas sederhana.

Berangkat dari evaluasi yang menekankan kemampuan model mendeteksi kelas minoritas (miskin ekstrem), hasil pada Tabel 2 menunjukkan bahwa pendekatan tree-based memberikan peningkatan kinerja yang bermakna dibanding baseline ekonometrika. XGBoost mencatat kemampuan diskriminasi tertinggi dengan AUC–ROC 0,93, diikuti Random Forest 0,90, sementara regresi logistik berada pada 0,84. Pola ini juga tercermin pada metrik yang lebih sensitif terhadap ketidakseimbangan kelas: F1-Score meningkat dari 0,42 (logistik) menjadi 0,56 (Random Forest) dan 0,61 (XGBoost), sedangkan Recall naik dari 0,55 menjadi 0,72 dan 0,78. Artinya, model boosting tidak hanya lebih tepat secara umum, tetapi lebih efektif mengurangi false negative yaitu rumah tangga miskin ekstrem yang luput terdeteksi yang dalam konteks kebijakan berisiko membuat kelompok paling rentan tidak terjangkau intervensi.

Secara substantif, temuan performa ini selaras dengan karakter kemiskinan ekstrem yang multidimensional dan cenderung membentuk pola non-linear: kombinasi keterbatasan modal manusia, rendahnya kualitas hunian, keterbatasan akses layanan dasar, serta kerentanan pasar kerja dapat saling memperkuat dan menciptakan jebakan (poverty trap). Dalam kerangka tersebut, regresi logistik tetap berguna untuk interpretabilitas, namun keterbatasannya muncul ketika risiko kemiskinan ekstrem terbentuk melalui interaksi variabel (misalnya pendidikan rendah bersamaan dengan aset minim, ketergantungan tinggi, dan akses air/bahan bakar yang buruk) yang sulit ditangkap oleh asumsi hubungan linier tunggal. Sebaliknya, Random Forest dan XGBoost secara alami menangkap interaksi dan non-linearitas, sehingga lebih adaptif untuk pemetaan risiko berbasis mikrodata seperti SUSENAS 2020–2024 yang diproses dengan pemisahan data latih-uji, tuning, serta evaluasi confusion matrix dan AUC–ROC.

Implikasi kebijakan dari hasil ini adalah pentingnya menempatkan Recall sebagai metrik prioritas ketika tujuan program adalah zero extreme poverty: lebih baik model “menjaring lebih banyak” kandidat miskin ekstrem (meski Precision tidak sempurna) daripada membiarkan kelompok paling rentan tidak teridentifikasi. Pada Tabel 2, Precision memang meningkat (0,34 → 0,46 → 0,50), tetapi tetap menunjukkan adanya trade-off klasik: semakin agresif menjaring kelompok miskin ekstrem, semakin tinggi pula potensi false positive. Karena itu, penggunaan model prediktif idealnya diposisikan sebagai screening tool untuk memperkuat penargetan (misalnya memprioritaskan verifikasi lapangan atau pemutakhiran basis data kesejahteraan),

bukan sebagai pengganti keputusan administratif tunggal. Di sisi metodologis, strategi penanganan ketidakseimbangan kelas melalui resampling pada data latih menjadi krusial agar model tidak bias ke kelas mayoritas, mengingat kemiskinan ekstrem merupakan kejadian yang relatif jarang.

Dari perspektif fenomena yang sedang terjadi di Jambi, pola prediksi juga masuk akal bila dikaitkan dengan dinamika makro dan sektoral yang telah dijelaskan dalam naskah: penurunan kemiskinan tajam pada 2017–2019 berkaitan dengan penguatan harga komoditas (karet dan sawit) yang menopang pendapatan rumah tangga pertanian, lalu terjadi lonjakan pada 2020–2021 akibat guncangan pandemi yang menekan aktivitas ekonomi dan sektor informal, serta kenaikan tipis kemiskinan pada September 2024 yang dikaitkan dengan tekanan inflasi dan harga kebutuhan pokok. Kondisi ini menguatkan argumen bahwa risiko kemiskinan ekstrem bersifat shock-sensitive—terutama pada rumah tangga yang bergantung pada sektor pertanian/perdesaan—dan menjelaskan mengapa model yang mampu menangkap perubahan pola serta interaksi variabel (seperti XGBoost) tampil lebih baik.

Sejalan dengan desain studi, analisis sektoral (subsampel pertanian dan non-pertanian) penting untuk membedakan determinan yang “universal” dan yang “spesifik sektor”. Secara teoritis, determinan universal yang konsisten muncul pada kedua sektor biasanya berkaitan dengan modal manusia dan kualitas hidup minimum—misalnya pendidikan kepala rumah tangga, struktur demografi (jumlah anggota dan dependency ratio), kualitas hunian (lantai/dinding), serta akses layanan dasar (air minum, bahan bakar memasak, jaminan kesehatan). Sementara itu, determinan spesifik sektor cenderung mengikuti mekanisme kerentanan utama: pada rumah tangga pertanian, volatilitas pendapatan dan produktivitas (dipengaruhi musim/komoditas) membuat kombinasi aset produktif-proksi (kualitas rumah/asset rumah tangga), struktur tanggungan, dan akses perlindungan sosial menjadi pengungkit utama untuk keluar dari kemiskinan ekstrem; sedangkan pada rumah tangga non-pertanian, risiko lebih kuat dipengaruhi ketidakstabilan pekerjaan informal, jam kerja, serta tekanan biaya hidup, sehingga variabel pasar kerja dan konektivitas (akses internet sebagai proksi akses informasi/pekerjaan) cenderung lebih dominan. Kerangka ini juga koheren dengan temuan spasial bahwa wilayah berdominasi pertanian (misalnya Tanjung Jabung Timur dan Barat) masih mencatat kemiskinan ekstrem lebih tinggi dibanding wilayah perkotaan.

## **KESIMPULAN**

Penelitian ini menunjukkan bahwa model machine learning, khususnya XGBoost, memberikan performa prediksi terbaik untuk klasifikasi status kemiskinan ekstrem rumah tangga di Provinsi Jambi berbasis mikrodata SUSENAS 2020–2024 (pooling), dengan AUC–ROC 0,93 dan Recall 0,78 pada data uji; Random Forest berada pada posisi kedua (AUC–ROC 0,90; Recall 0,72), sedangkan regresi logistik sebagai baseline masih kompetitif namun lebih rendah (AUC–ROC 0,84; Recall 0,55). Temuan ini menegaskan bahwa pendekatan tree-based lebih sesuai untuk fenomena kemiskinan ekstrem yang non-linear dan dipengaruhi interaksi variabel, sementara regresi logistik tetap bernilai untuk kebutuhan interpretasi kebijakan. Dalam konteks target penghapusan kemiskinan ekstrem, metrik Recall menjadi kunci karena menekan risiko rumah tangga miskin ekstrem yang tidak teridentifikasi oleh sistem

penargetan.

## DAFTAR PUSTAKA

- Aron, D. V., Castaneda Aguilar, R. A., Diaz-Bonilla, C., Farfan Betran, M. G., Foster, E. M., Fujs, T. H. M. J., Jolliffe, D., Krishnan, N., Lakner, C., Lara Ibarra, G., Mahler, D. G., Moreno Herrera, L., Nguyen, M. C., Sanchez Castro, D. M., Tetteh-Baah, S. K., Viveros Mendoza, M. C., Wu, H., & Yonzan, N. (2023). September 2023 Update to the Poverty and Inequality Platform (PIP): What's New. September 2023 Update to the Poverty and Inequality Platform (PIP): What's New, March. <https://doi.org/10.1596/40455>
- BPS. (2025). Profil Kemiskinan di Jambi, September 2024. 09.
- Brown, C., Ravallion, M., & van de Walle, D. (2018). A poor means test? Econometric targeting in Africa. *Journal of Development Economics*, 134, 109–124. <https://doi.org/10.1016/j.jdeveco.2018.05.004>
- Christiaensen, L. U. C. (2013). Poverty Reduction During the Rural – Urban Transformation – The Role of the Missing Middle. *WORLD DEVELOPMENT*, xx. <https://doi.org/10.1016/j.worlddev.2013.10.002>
- Faharuddin, F., Kependudukan, B., Keluarga, D., Nasional, B., Endrawati, D., & Indonesia, S. (2022). Determinants of working poverty in Indonesia. April. <https://doi.org/10.1108/JED-09-2021-0151>
- Fatmah, F. (2024). Factors associated with food security in Depok City, Indonesia during the COVID-19 pandemic: a cross-sectional study. *Frontiers in Sustainable Food Systems*, 8(April), 1–10. <https://doi.org/10.3389/fsufs.2024.1327887>
- Fosu, A. K. (2017). Growth, inequality, and poverty reduction in developing countries: Recent global evidence. *Research in Economics*, 71(2), 306–336. <https://doi.org/10.1016/j.rie.2016.05.005>
- International Labour Office. (1976). *Employment, Growth and Basic Needs: A One-World Problem*. Report of the Director-General of the International Labour Office.
- Kuznets, S. (1955). Kuznets1955. In *The American Economic Review* (Vol. 1, p. 30). <https://www.jstor.org/stable/1811581>
- Lagakos, D., & Waugh, M. E. (2012). Selection , Agriculture , and Cross-Country Productivity Differences.
- Lewis, W. A. (1954). “ Economic Development with Unlimited Supplies of Labour ”\*.
- Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 2017-Decem(Section 2), 4766–4775.
- Mcmillan, M., Dc, W., & Rodrik, D. (2014). Globalization , Structural Change , and Productivity Growth , with an Update on Africa. 63, 11–32.

- <https://doi.org/10.1016/j.worlddev.2013.10.012>
- Michael P. Todaro, S. C. S. (2015). *Economic Development* (12th Editi). Pearson Education Limited.
- OECD. (2019). *The Path to Becoming a Data-Driven Public Sector* (2019th ed.).
- Olivia, S., & Gibson, J. (2013). Using Engel curves to measure CPI bias for Indonesia. *Bulletin of Indonesian Economic Studies*, 49(1), 85–101. <https://doi.org/10.1080/00074918.2013.772942>
- Piketty, T. (2014). *Capitalism in the 21st Century*. In *Capitalism in the 21st Century*. <https://doi.org/10.2307/j.ctv33mgb41>
- Piketty, T. (2021). Capital and ideology: A global perspective on inequality regimes. *British Journal of Sociology*, 72(1), 139–150. <https://doi.org/10.1111/1468-4446.12836>
- Purwono, R., Wardana, W. W., Haryanto, T., & Khoerul Mubin, M. (2021). Poverty dynamics in Indonesia: empirical evidence from three main approaches. *World Development Perspectives*, 23, 100346. <https://doi.org/10.1016/j.wdp.2021.100346>
- Putri, S. R., Wijayanto, A. W., & Pramana, S. (2023). Multi-source satellite imagery and point of interest data for poverty mapping in East Java, Indonesia: Machine learning and deep learning approaches. *Remote Sensing Applications: Society and Environment*, 29, 100889. <https://doi.org/https://doi.org/10.1016/j.rsase.2022.100889>
- Rezaei, S., Karimi, M., Soltani, S., Barfar, E., Gharehghani, M. A. M., Badakhshan, A., Badiiee, N., Pakdaman, M., & Brown, H. (2024). Household financial burden associated with out-of-pocket payments for healthcare in Iran: insights from a cross-sectional survey. *BMC Health Services Research*, 24(1). <https://doi.org/10.1186/s12913-024-11477-z>
- Rodrik, D., & Rodrik, D. (2015). No Title.
- Rogerson, R., & Rogerson, R. (2013). No Title.
- Ruja, I. N., Sumarmi, & Idris. (2024). Programs, Opportunities, and Challenges in Poverty Reduction: A Systematic Review. *SAGE Open*, 14(2), 1–13. <https://doi.org/10.1177/21582440241256242>
- T. Hastie, R. Tibshirani, and J. F. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. New York, NY, USA: Springer, 2009. *The Mathematical Intelligencer*, 27(2), 83–85.
- Tohari, A., Parsons, C., & Rammohan, A. (2019). Targeting poverty under complementarities : Evidence from Indonesia ' s uni fi ed targeting system ☆. *Journal of Development Economics*, 140(April 2018), 127–144. <https://doi.org/10.1016/j.jdeveco.2019.06.002>
- Yuliana, R., Harianto, N., Hartoyo, S., & Firdaus, M. (2019). Dampak Perubahan Harga Pangan terhadap Tingkat Kesejahteraan Rumah Tangga di Indonesia.

Jurnal Agro Ekonomi, 37(1), 25. <https://doi.org/10.21082/jae.v37n1.2019.25-45>  
Zevaya, F., & Asfahani, F. R. (2025). Evidence-based strategies for extreme poverty reduction : Identifying priority areas in Jambi Province , Indonesia. 13(2), 95–108. <https://doi.org/10.22437/ppd.v13i2.38125>