JITET (Jurnal Informatika dan Teknik Elektro Terapan)
Vol. 14 No. 3, pISSN: 2303-0577 eISSN: 2830-7062 http://dx.doi.org/10.23960/jitet.v14i3.10699

# PREDIKSI INDEKS KEPARAHAN KEMISKINAN KABUPATEN/KOTA DI PROVINSI JAWA BARAT MENGGUNAKAN ALGORITMA RANDOM FOREST

**Ririn Rusmayanti[1*], Aries Suharso[2], Chaerur Rozikin[3]**

[1,2,3]Universitas Singaperbangsa Karawang, Jl.HS.RonggoWaluyo, Puseurjaya, Telukjambe Timur, Karawang, Jawa Barat 41361, Telp. (0267) 641177

**Keywords:**
Indeks Keparahan Kemiskinan; Random Forest; CRISP-DM; Prediksi; Jawa Barat.

**Corespondent Email:**
ririnrusmayanti35@gmail.com

**Abstrak.** Kemiskinan merupakan masalah multidimensi yang tidak hanya diukur dari persentase penduduk miskin, tetapi juga ketimpangan pengeluaran di antara mereka, yang direpresentasikan oleh Indeks Keparahan Kemiskinan (P2). Penelitian ini bertujuan untuk memprediksi nilai P2 pada 27 kabupaten/kota di Provinsi Jawa Barat menggunakan algoritma Random Forest. Data yang digunakan merupakan data sekunder dari BPS Provinsi Jawa Barat periode 2015–2025 dengan 297 observasi. Tujuh variabel prediktor sosial ekonomi digunakan, termasuk persentase penduduk miskin dan PDRB. Metodologi yang digunakan adalah CRISP-DM. Proses pemodelan melibatkan hyperparameter tuning menggunakan GridSearchCV. Model terbaik (RF_Tuning_n500_depth5_feat7_split2_leaf6) menghasilkan nilai Mean Absolute Error (MAE) sebesar 0,0690, Root Mean Squared Error (RMSE) sebesar 0,0918, dan Koefisien Determinasi ($R^2$) sebesar 0,6542. Analisis feature importance menunjukkan bahwa persentase penduduk miskin memberikan kontribusi terbesar (85,05%) terhadap model. Hasil penelitian menunjukkan bahwa algoritma Random Forest terbukti efektif dalam memprediksi tingkat keparahan kemiskinan

***Abstract.*** *Poverty is a multidimensional problem measured not only by the percentage of the poor population but also by the expenditure inequality among them, represented by the Poverty Severity Index (P2). This study aims to predict the P2 value in 27 regencies/cities in West Java Province using the Random Forest algorithm. The data used is secondary data from the West Java Province BPS for the 2015–2025 period with 297 observations. Seven socioeconomic predictor variables were used. The methodology used is CRISP-DM. The modeling process involves hyperparameter tuning using GridSearchCV. The best model produced a Mean Absolute Error (MAE) of 0.0690, a Root Mean Squared Error (RMSE) of 0.0918, and a Coefficient of Determination ($R^2$) of 0.6542. Feature importance analysis shows that the percentage of the poor population provides the most significant contribution (85.05%) to the prediction.*



## 1. PENDAHULUAN

Kemiskinan merupakan salah satu permasalahan pembangunan yang hingga kini menjadi perhatian utama berbagai negara, termasuk Indonesia. Kemiskinan tidak hanya ditandai oleh rendahnya pendapatan atau pengeluaran, tetapi juga berkaitan dengan keterbatasan pemenuhan kebutuhan dasar seperti pendidikan, kesehatan, pangan, dan tempat tinggal yang layak [1]. Berdasarkan data Badan Pusat Statistik (BPS), jumlah penduduk miskin di Provinsi Jawa Barat pada September 2025 tercatat 3,55 juta jiwa (6,78%), menurun dibandingkan September 2024 sebesar 3,67 juta jiwa (7,08%). Meskipun demikian, penurunan tersebut belum sepenuhnya mencerminkan membaiknya kondisi kemiskinan, karena indikator jumlah penduduk miskin hanya menggambarkan proporsi penduduk di bawah garis kemiskinan tanpa memperhitungkan

kedalaman maupun ketimpangan pengeluaran antarpenduduk miskin [2].

BPS mengukur tingkat kemiskinan menggunakan pendekatan Foster-Greer-Thorbecke (FGT), yang terdiri atas Persentase Penduduk Miskin (P0), Indeks Kedalaman Kemiskinan (P1), dan Indeks Keparahan Kemiskinan (P2) [2]. Indeks Keparahan Kemiskinan memberikan bobot lebih besar pada penduduk dengan pengeluaran jauh di bawah garis kemiskinan sehingga mencerminkan ketimpangan pengeluaran di antara penduduk miskin; semakin tinggi nilai P2, semakin parah kondisi kemiskinan suatu wilayah. Dengan demikian, P2 menjadi indikator pelengkap yang penting bagi P0 agar analisis kemiskinan dapat dilakukan secara lebih komprehensif [3]. Nilai P2 Provinsi Jawa Barat menunjukkan tren menurun secara umum sepanjang 2015–2025, meski sempat melonjak tajam pada 2021 (dari 0,23 menjadi 0,38) akibat pandemi COVID-19, menegaskan bahwa P2 bersifat dinamis dan dipengaruhi banyak faktor sosial ekonomi.

Kondisi tersebut menuntut pendekatan yang mampu menghasilkan nilai prediksi P2 berdasarkan indikator sosial ekonomi. Pendekatan berbasis machine learning semakin banyak dimanfaatkan untuk keperluan ini. Algoritma Random Forest dipilih pada penelitian ini karena kemampuannya memodelkan hubungan nonlinier antarvariabel, menangani data dengan karakteristik kompleks, relatif tahan terhadap overfitting, serta mampu memberikan informasi kontribusi tiap variabel melalui feature importance [4].

Algoritma Random Forest telah banyak diterapkan dalam kajian kemiskinan maupun bidang prediktif lain. Studi Khalik dan Arifin, membandingkan Decision Tree, K-Nearest Neighbor, Naive Bayes, Neural Network, dan Random Forest untuk mengklasifikasikan Indeks Kedalaman Kemiskinan (P1) di Sulawesi Selatan [5], sedangkan studi Heryati dan Setiawan menggunakan Random Forest untuk memprediksi tingkat kemiskinan di Sumatera Selatan dengan performa terbaik ($R^2$ = 0,7244; MAE = 0,2489) dibandingkan Linear Regression, XGBoost, dan Neural Network [6]. Penelitian Valentika menerapkan Random Forest untuk membangun interval prediksi pada data kemiskinan 514 kabupaten/kota di Indonesia [7]. Penerapan Random Forest juga dijumpai pada berbagai kajian prediktif mengenai prediksi jumlah santri baru [8] dan analisis sentimen berbasis Random Forest classifier, yang menunjukkan fleksibilitas algoritma ini pada berbagai domain data tabular [9].

Berdasarkan tinjauan tersebut, sebagian besar penelitian masih berfokus pada prediksi tingkat kemiskinan secara umum atau klasifikasi P0/P1, sedangkan penelitian yang secara khusus menghitung nilai prediksi Indeks Keparahan Kemiskinan (P2) pada tingkat kabupaten/kota di Provinsi Jawa Barat menggunakan Random Forest masih sangat terbatas. Oleh karena itu, penelitian ini bertujuan menghitung nilai prediksi Indeks Keparahan Kemiskinan pada 27 kabupaten/kota di Provinsi Jawa Barat menggunakan algoritma Random Forest berdasarkan tujuh variabel sosial ekonomi, dan mengevaluasi performa algoritma tersebut menggunakan metrik MAE, RMSE, dan $R^2$, serta mengidentifikasi variabel yang paling berkontribusi melalui analisis feature importance.

## 2. TINJAUAN PUSTAKA

Indeks Keparahan Kemiskinan (P2) merupakan instrumen pengukuran yang krusial untuk melengkapi indikator kemiskinan dasar, karena indeks ini tidak sekadar menghitung jumlah penduduk miskin, melainkan mengukur tingkat ketimpangan pengeluaran di antara penduduk miskin itu sendiri. Dalam praktiknya, suatu wilayah mungkin memiliki persentase penduduk miskin yang tinggi namun dengan kesenjangan pengeluaran yang relatif merata, atau sebaliknya, jumlah penduduk miskin sedikit tetapi ketimpangannya sangat tajam. Oleh karena itu, P2 yang merupakan bagian dari keluarga indeks *Foster-Greer-Thorbecke* (FGT) hadir untuk memberikan gambaran yang lebih utuh dan sensitif mengenai distribusi kesejahteraan pada kelompok masyarakat paling bawah [10]. Semakin besar disparitas pengeluaran di antara sesama penduduk miskin dalam suatu wilayah, maka semakin tinggi pula nilai P2 yang dihasilkan.

Secara matematis, nilai P2 diperoleh melalui agregasi kuadrat dari rasio defisit pengeluaran setiap individu miskin terhadap garis kemiskinan. Operasi pemangkatan dua ini menjadi karakteristik fundamental karena

secara otomatis memberikan bobot yang semakin besar bagi individu yang tingkat pengeluarannya berada sangat jauh di bawah garis kemiskinan. Selain itu, perhitungan ini menggunakan jumlah seluruh penduduk (total populasi) sebagai penyebut, bukan sekadar jumlah penduduk miskin. Artinya, nilai indeks ini dipengaruhi secara simultan oleh besarnya ketimpangan di antara penduduk miskin sekaligus proporsi penduduk miskin tersebut terhadap total populasi suatu wilayah [3]. P2 dirumuskan sebagai berikut:

$$P_2 = \frac{1}{n}\sum_{i=1}^{q}\left(\frac{z - y_i}{z}\right)^2$$

Keterangan:

- $P_2$= Indeks Keparahan Kemiskinan
- $n$= jumlah seluruh penduduk
- $q$= jumlah penduduk miskin
- $z$= garis kemiskinan
- $y_i$= pengeluaran per kapita penduduk miskin ke-$i$

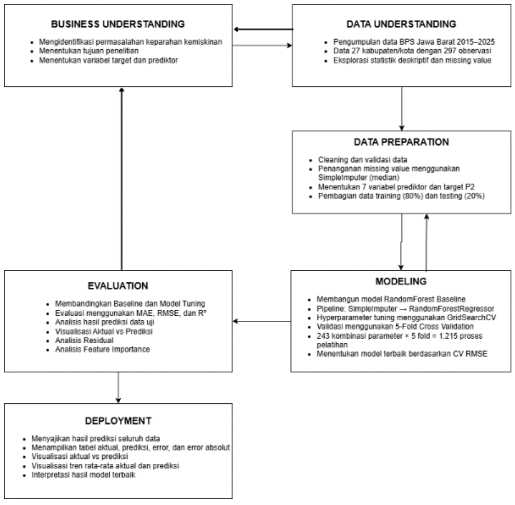
Random Forest merupakan algoritma *ensemble learning* yang beroperasi dengan membangun sekumpulan pohon keputusan (*decision tree*) secara independen selama proses pelatihan. Pembangunan algoritma ini dilandaskan pada dua mekanisme utama, yaitu *bootstrap aggregating* (*bagging*) dan *random feature selection*. Pada proses *bagging*, setiap pohon dilatih menggunakan *bootstrap sample*, yakni sampel data yang diambil secara acak dengan pengembalian (*sampling with replacement*) dari *dataset* utama. Selanjutnya, pada setiap pemisahan *node* (*split*), algoritma hanya memilih sebagian fitur secara acak untuk menentukan pemisahan terbaik.

Pada prediksi menggunakan Random Forest hasil akhir didapatkan dengan menghitung rata-rata (*averaging*) dari seluruh keluaran prediksi masing-masing pohon keputusan. Penggabungan banyak pohon yang saling tidak berkorelasi ini mampu meminimalisir variansi prediksi, meningkatkan kemampuan generalisasi model, dan menekan risiko *overfitting* secara signifikan dibandingkan penggunaan satu pohon keputusan tunggal. Selain kemampuan prediktifnya, algoritma ini juga menyediakan fitur pengukuran tingkat kepentingan variabel (*feature importance*) berbasis penurunan *Mean Squared Error* (MSE). Fitur ini secara interpretatif mampu mengukur dan mengurutkan variabel prediktor mana yang memberikan kontribusi paling besar dalam meminimalisir tingkat kesalahan selama proses pembangunan model.

## 3. METODE PENELITIAN

Pada penelitian ini menggunakan data sekunder berupa data panel Indeks Keparahan Kemiskinan (P2) 27 kabupaten/kota di Provinsi Jawa Barat periode 2015–2025 (297 observasi) yang diperoleh dari Badan Pusat Statistik Provinsi Jawa Barat melalui pengajuan permohonan data resmi. Indeks Keparahan Kemiskinan digunakan sebagai variabel target, sedangkan tujuh variabel lain digunakan sebagai prediktor, yaitu Tingkat Partisipasi Angkatan Kerja (%), Pengeluaran per Kapita (ribu Rp/orang/tahun), PDRB (ribu Rp), Rata-rata Lama Sekolah (tahun), Usia Harapan Hidup (tahun), Persentase Penduduk Miskin (%), dan Garis Kemiskinan (Rp/kapita/bulan). Ketujuh variabel tersebut dipilih berdasarkan keterkaitan teoritis terhadap P2 serta ketersediaan data yang lengkap pada seluruh kabupaten/kota.

Metode penelitian yang digunakan dalam penelitian ini adalah CRISP-DM (*Cross-Industry Standard Process for Data Mining*) yang terdiri atas enam tahapan, yaitu *business understanding, data, data preparation, modeling, evaluation,* dan *deployment.* Agar dapat lebih mudah dipahami, alur penelitian disajikan pada Gambar 1.




**Gambar 1.** Alur Penelitian

### *3.1. Business Understanding*

Business understanding merupakan tahap untuk memahami konteks permasalahan dan menetapkan tujuan penelitian. Pada penelitian ini, tahap tersebut digunakan untuk merumuskan tujuan penelitian, yaitu memprediksi Indeks Keparahan Kemiskinan menggunakan algoritma Random Forest serta mengidentifikasi variabel sosial ekonomi yang paling berpengaruh [11].

### *3.2. Data Understanding*

Data understanding merupakan tahap untuk mengumpulkan dan memahami karakteristik data yang akan digunakan. Pada penelitian ini, tahap tersebut digunakan untuk memeriksa struktur, kelengkapan, dan sebaran data BPS Provinsi Jawa Barat sebelum data diolah lebih lanjut.

### *3.3. Data Preparation*

Data preparation merupakan tahap untuk mempersiapkan data agar siap digunakan dalam pemodelan. Pada penelitian ini, tahap tersebut digunakan untuk menentukan variabel prediktor dan target, membagi data menjadi data latih dan data uji, serta menangani nilai yang hilang melalui Pipeline dengan SimpleImputer.

### *3.4. Modeling*

Modeling merupakan tahap untuk membangun model prediktif dari data yang telah disiapkan. Pada penelitian ini, tahap tersebut digunakan untuk membangun model RandomForestRegr essor sebagai baseline, kemudian melakukan hyperparameter tuning menggunakan GridSearchCV dengan 5-fold cross-validation untuk memperoleh kombinasi parameter terbaik.

### *3.5. Evaluation*

Evaluation merupakan tahap untuk menilai performa model yang telah dibangun. Pada penelitian ini, tahap tersebut digunakan untuk mengukur performa model menggunakan MAE, RMSE, dan $R^2$, serta menganalisis feature importance dan residual hasil prediksi.

### *3.6. Deployment*

*Deployment* merupakan tahap untuk menerapkan model pada keseluruhan data dan menyajikan hasil akhir. Pada penelitian ini, tahap tersebut digunakan untuk menghasilkan prediksi Indeks Keparahan Kemiskinan pada seluruh 297 observasi serta memvisualisasikan tren aktual dan prediksi per tahun.

## 4. HASIL DAN PEMBAHASAN

### *4.1. Business Understanding*

Berdasarkan eksplorasi awal data BPS Provinsi Jawa Barat, ditemukan bahwa Indeks Keparahan Kemiskinan antarkabupaten/kota menunjukkan pola yang fluktuatif dan heterogen, sehingga tidak dapat diprediksi hanya dengan pendekatan sederhana dan memerlukan model prediktif berbasis machine learning. Berdasarkan hal tersebut, tujuan penelitian ditetapkan yaitu menerapkan model prediksi Random Forest untuk menghitung nilai Indeks Keparahan Kemiskinan pada 27 kabupaten/kota di Provinsi Jawa Barat, mengevaluasi performa model menggunakan metrik MAE, RMSE, dan $R^2$, serta mengidentifikasi variabel sosial ekonomi yang paling berpengaruh melalui analisis feature importance.

### *4.2. Data Understanding*

Dataset yang diperoleh dari BPS Provinsi Jawa Barat diimpor ke lingkungan Python menggunakan pustaka Pandas, kemudian diperiksa strukturnya meliputi jumlah baris dan kolom (297 baris, 10 kolom), nama dan tipe data setiap variabel, jumlah kabupaten/kota (27), serta rentang tahun pengamatan (2015-2025). Pemeriksaan missing value dan data duplikat tidak menemukan adanya nilai yang hilang maupun baris data ganda, sehingga dataset dinyatakan layak digunakan untuk tahap pemodelan. Selanjutnya dilakukan analisis statistik deskriptif untuk memperoleh gambaran nilai minimum, maksimum, rata-rata, dan sebaran tiap variabel, serta analisis korelasi antarvariabel untuk mengidentifikasi variabel yang memiliki keterkaitan paling kuat terhadap variabel target.

### *4.3. Data Preparation*

Tahap ini diawali dengan data cleaning, yaitu memastikan seluruh variabel numerik memiliki tipe data yang sesuai, dilanjutkan dengan penentuan variabel prediktor (X) dan variabel target (Y); Indeks Keparahan Kemiskinan digunakan sebagai target, sedangkan ketujuh variabel lainnya digunakan sebagai prediktor. Metadata berupa nama kabupaten/kota dan tahun turut disiapkan untuk keperluan penyajian hasil prediksi pada tahap *deployment*. Dataset kemudian dibagi menjadi data latih (80%, 237 observasi) dan data uji (20%, 60 observasi) menggunakan metode train-test split. Untuk menjaga konsistensi

praproses serta mengurangi risiko kebocoran data (data *leakage*) antara data latih dan data uji, penelitian ini menggunakan Pipeline yang terdiri atas SimpleImputer dengan strategi median, sehingga proses imputasi nilai yang hilang dapat dilakukan secara konsisten dan otomatis pada kedua subset data.

### *4.4. Modeling*

Tahap modeling dilakukan melalui dua tahap. Tahap pertama membentuk model RandomForest_Baseline dengan n_estimators=500, max_depth tidak dibatasi, dan seluruh tujuh fitur digunakan sebagai max_features. Tahap kedua melakukan hyperparameter tuning menggunakan GridSearchCV dengan 5-fold cross-validation terhadap parameter pada Tabel 1, mengikuti pendekatan yang lazim digunakan pada optimasi Random Forest berbasis grid search, menghasilkan total 3×3×3×3×3 = 243 kombinasi (1.215 fits).

**Tabel 1 .** *Grid Hyperparameter* yang Diuji

| Parameter | Nilai Diuji |
|---|---|
| n_estimators | 300, 500, 700 |
| max_depth | 5, 10, 15 |
| max_features | 3, 5, 7 |
| min_samples_split | 2, 5, 10 |
| min_samples_leaf | 2, 4, 6 |

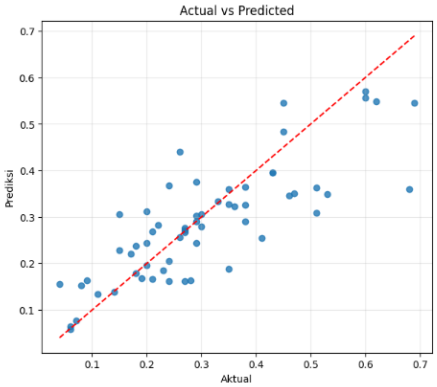
Model RandomForest_Baseline menghasilkan MAE sebesar 0,0665, RMSE sebesar 0,0921, dan $R^2$ sebesar 0,6527. Kombinasi terbaik hasil GridSearchCV adalah n_estimators=500, max_depth=5, max_features=7, min_samples_split=2, dan min_samples_leaf=6, dengan RMSE cross-validation sebesar 0,0954. Ketika dievaluasi pada data uji, model RF_Tuning_n500_depth5_feat7_split2_leaf6 ini menghasilkan MAE sebesar 0,0690, RMSE sebesar 0,0919, dan $R^2$ sebesar 0,6542.

**Tabel 2.** Perbandingan Model Baseline dan Tuning Terbaik

| Model | MAE | RMSE | $R^2$ |
|---|---|---|---|
| RandomForest_Baseline | 0,0665 | 0,0921 | 0,6527 |
| RF_Tuning_n500_depth5_feat7_split2_leaf6 | 0,0690 | 0,0919 | 0,6542 |

Meskipun MAE model tuning sedikit lebih besar dibandingkan baseline, model tuning tetap dipilih sebagai model akhir karena RMSE lebih rendah dan $R^2$ lebih tinggi. RMSE dijadikan pertimbangan utama karena digunakan sebagai fungsi scoring selama GridSearchCV dan lebih sensitif terhadap kesalahan prediksi besar, yang penting dalam konteks P2 karena kesalahan besar pada suatu wilayah berdampak lebih signifikan dibandingkan penurunan kecil pada rata-rata kesalahan absolut. Nilai $R^2$ sebesar 0,654 menunjukkan bahwa sekitar 65,4% variasi Indeks Keparahan Kemiskinan pada data uji dapat dijelaskan oleh ketujuh variabel prediktor, sedangkan 34,6% sisanya dipengaruhi faktor lain di luar model, seperti kondisi ekonomi makro maupun kebijakan pemerintah.

### *4.5. Evaluation*

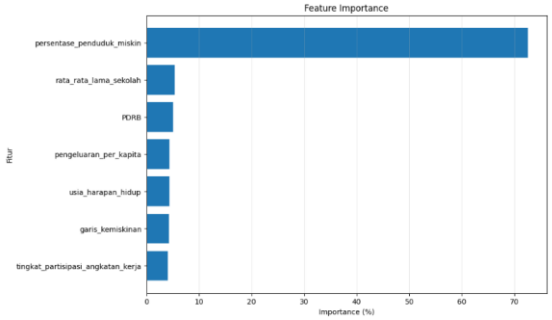



**Gambar 2.** Perbandingan Nilai Aktual dan Prediksi pada Data Uji

Gambar 2 menunjukkan hubungan antara nilai aktual dan nilai prediksi pada data uji. Sebagian besar titik berada di sekitar garis diagonal, mengindikasikan bahwa model mampu mengikuti pola umum data dengan baik, meskipun beberapa observasi seperti Kota Cirebon (2021) menunjukkan selisih prediksi yang relatif besar (error 0,3156, sekitar 4,6 kali MAE), menunjukkan model cenderung underestimate pada nilai P2 yang sangat tinggi. Sebaliknya, pada Kabupaten Garut (2020), model overestimate dengan error -0,176. Pola ini konsisten dengan analisis residual, di mana residual tersebar di sekitar nol tanpa pola sistematis yang jelas sehingga tidak terdapat bias signifikan, meskipun beberapa residual besar tetap muncul pada nilai P2 yang ekstrem.

**Tabel 3.** *Feature Importance*

| Feature | Importance (%) |
|---|---|
| Persentase penduduk miskin | 85,05 |
| Pengeluaran per kapita | 3,12 |
| PDRB | 3,05 |
| Rata-rata lama sekolah | 2,55 |
| Usia harapan hidup | 2,41 |
| Garis kemiskinan | 2,15 |
| Tingkat partisipasi angkatan kerja | 1,68 |

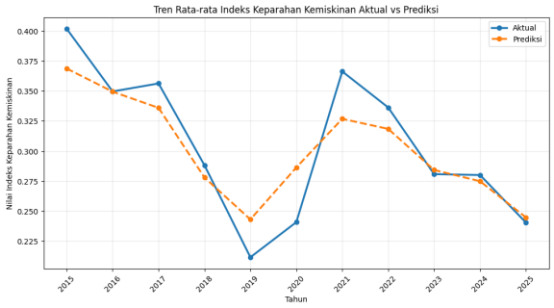



**Gambar 3.** Grafik *Feature Importance*

Analisis *feature importance* (Tabel 3, Gambar 3) menunjukkan bahwa variabel persentase penduduk miskin memiliki kontribusi dominan sebesar 85,05%, jauh melampaui variabel lainnya: pengeluaran per kapita (3,12%), PDRB (3,05%), rata-rata lama sekolah (2,55%), usia harapan hidup (2,41%), garis kemiskinan (2,15%), dan tingkat partisipasi angkatan kerja (1,68%). Dominasi ini konsisten dengan hasil analisis korelasi pada tahap *data understanding*, yang menunjukkan hubungan terkuat antara persentase penduduk miskin dan P2, sehingga memperkuat variabel tersebut sebagai prediktor utama dalam model. Indikator terkait kondisi penduduk miskin secara langsung merupakan prediktor utama dalam model Random Forest untuk kajian kemiskinan wilayah.

### *4.6. Deployment*

Model terbaik diterapkan untuk memprediksi P2 pada seluruh 297 observasi. Visualisasi tren rata-rata tahunan (Gambar 4) menunjukkan garis prediksi mengikuti pola garis aktual, termasuk penurunan pada 2018–2019, kenaikan pada 2021 akibat pandemi, serta penurunan kembali setelah 2022, meskipun pada beberapa tahun (2019–2020 dan 2021–2022) masih terdapat selisih yang tidak sepenuhnya sama.




**Gambar 4.** Tren Rata-Rata Indeks Keparahan Kemiskinan Aktual dan Prediksi per Tahun

Kesesuaian arah tren antara nilai aktual dan prediksi, termasuk pada momen pandemi COVID-19 tahun 2021, menunjukkan bahwa model tidak hanya menghafal pola data latih, tetapi juga mampu menangkap dinamika perubahan kondisi sosial ekonomi yang memengaruhi keparahan kemiskinan dari waktu ke waktu. Selisih yang masih muncul pada periode 2019-2020 dan 2021-2022 mengindikasikan bahwa model cenderung merespons perubahan tren dengan sedikit keterlambatan, sebuah karakteristik yang lazim pada model berbasis rata-rata banyak pohon seperti Random Forest ketika menghadapi perubahan yang tiba-tiba. Dengan demikian, hasil deployment ini menegaskan bahwa model cukup andal untuk menggambarkan pola keparahan kemiskinan secara agregat, meskipun perlu kehati-hatian saat digunakan untuk mendeteksi perubahan mendadak pada tahun tertentu.

### *Perbandingan dengan Penelitian Terdahulu*

Nilai $R^2$ sebesar 0,654 pada penelitian ini lebih rendah dibandingkan dengan penelitian Heryati dkk., yang memperoleh $R^2$ sebesar 0,7244 dalam memprediksi tingkat kemiskinan di Sumatera Selatan. Perbedaan ini diduga terkait dengan karakteristik variabel target: penelitian ini memprediksi Indeks Keparahan Kemiskinan (P2) yang mengukur ketimpangan pengeluaran antarpenduduk miskin dan cenderung lebih fluktuatif antarwaktu, sedangkan penelitian tersebut memprediksi tingkat kemiskinan secara umum yang polanya relatif lebih stabil [6].

Temuan *feature importance* pada penelitian ini, yaitu dominannya persentase penduduk

miskin (85,05%), juga sejalan dengan penelitian Handayani dan Qutub yang menemukan bahwa indikator kondisi penduduk miskin secara langsung merupakan prediktor utama dalam model Random Forest untuk kajian kemiskinan wilayah, serta memperkuat yang menekankan pentingnya interval prediksi berbasis Random Forest dalam menangkap ketidakpastian data kemiskinan pada level kewilayahan [12].

Hasil evaluasi pada penelitian ini juga sejalan dengan temuan Mushagalusa, yang menyatakan bahwa performa Random Forest dipengaruhi oleh karakteristik dataset seperti ukuran sampel, jumlah atribut, dan variasi data [13], serta penelitian Han dkk., yang menunjukkan bahwa penerapan Random Forest pada dataset berukuran relatif kecil menghadapi tantangan berupa keterbatasan observasi dibandingkan jumlah variabel, sehingga hyperparameter tuning tidak selalu memberikan peningkatan performa signifikan [14]. Temuan tersebut konsisten dengan hasil penelitian ini, di mana tuning hanya meningkatkan $R^2$ dari 0,6527 menjadi 0,6542 dan menurunkan RMSE dari 0,0921 menjadi 0,0919, mengindikasikan bahwa dataset panel berukuran 297 observasi telah mendekati batas kemampuan generalisasi Random Forest tanpa penambahan data atau variabel baru. Analisis pola residual pada penelitian ini turut mengacu pada kerangka evaluasi kesalahan prediksi model regresi yang diuraikan oleh Verma [15].

### *Implikasi Penelitian*

Secara teoretis, dominannya kontribusi variabel persentase penduduk miskin terhadap prediksi P2 memperkuat kerangka Foster-Greer-Thorbecke, di mana P2 merupakan turunan langsung dari proporsi dan kedalaman pengeluaran penduduk miskin, sehingga secara empiris menegaskan keterkaitan antarindikator FGT yang selama ini dijelaskan secara teoretis. Temuan ini juga memperkaya literatur penerapan Random Forest pada data panel berukuran kecil, dengan menunjukkan bahwa kontribusi hyperparameter tuning terhadap peningkatan performa menjadi terbatas ketika jumlah observasi relatif kecil dibandingkan jumlah kombinasi parameter yang diuji. Secara praktis, model prediksi ini dapat dimanfaatkan oleh pemerintah daerah maupun BPS sebagai alat bantu deteksi dini wilayah dengan risiko keparahan kemiskinan tinggi berdasarkan indikator sosial ekonomi yang relatif mudah dipantau, khususnya persentase penduduk miskin, sehingga kebijakan pengentasan kemiskinan dapat dirumuskan secara lebih tepat sasaran, baik dari sisi wilayah maupun periode waktu intervensi.

## 5. KESIMPULAN

Berdasarkan hasil penelitian dan pembahasan yang telah dilakukan, dapat ditarik beberapa kesimpulan sebagai berikut:

a. Algoritma Random Forest berhasil digunakan untuk memprediksi Indeks Keparahan Kemiskinan pada 27 kabupaten/kota di Provinsi Jawa Barat menggunakan tujuh variabel sosial ekonomi melalui pendekatan CRISP-DM dan Pipeline. Model terbaik hasil *hyperparameter tuning* memperoleh nilai MAE sebesar 0,0690, RMSE sebesar 0,092, dan $R^2$ sebesar 0,654, sehingga mampu menjelaskan sekitar 65,4% variasi Indeks Keparahan Kemiskinan pada data uji.
b. Analisis *feature importance* menunjukkan bahwa persentase penduduk miskin merupakan variabel yang paling berkontribusi terhadap prediksi (85,05%), sehingga menjadi indikator utama yang perlu diperhatikan dalam penyusunan kebijakan pengentasan kemiskinan berbasis wilayah.
c. Kelebihan penelitian ini adalah Random Forest mampu memberikan performa prediksi yang cukup baik serta mengidentifikasi tingkat kepentingan setiap variabel melalui analisis *feature importance*.
d. Keterbatasan penelitian ini terletak pada jumlah observasi yang relatif terbatas dan penggunaan variabel prediktor yang belum mencakup seluruh faktor yang memengaruhi Indeks Keparahan Kemiskinan.
e. Penelitian selanjutnya disarankan menggunakan dataset yang lebih besar, menambahkan variabel prediktor yang relevan, serta membandingkan Random Forest dengan algoritma *ensemble* lain untuk meningkatkan performa prediksi.

## UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada seluruh pihak yang telah memberikan dukungan, bantuan, dan kontribusi selama pelaksanaan penelitian ini sehingga penelitian dapat diselesaikan dengan baik.

## DAFTAR PUSTAKA

[1] P. Indah, S. I. Retno, G. Tian, D. Ari, A. Dwi, And H. M. Al, "Analisis Pengaruh Tingkat Pengangguran Terbuka Dan Indeks Pembangunan Manusia Terhadap Kemiskinan Di Provinsi Jawa Tengah," *J. Saintika Unpam J. Sains Dan Mat. Unpam*, Vol. 3, No. 1, Pp. 81–88, 2020.

[2] Bps, "Profil Kemiskinan Di Jawa Barat Maret 2025," 2025. [Online]. Available: Https://Jabar.Bps.Go.Id [Accessed: July. 20, 2026]

[3] D. Septiadi And M. Nursan, "Pengentasan Kemiskinan Indonesia: Analisis Indikator Makroekonomi Dan Kebijakan Pertanian," *J. Hexagro*, Vol. 4, No. 1, Pp. 1–14, 2020, Doi: 10.36423/Hexagro.V4i1.371.

[4] L. Breiman, "Random Forests," *2001 Kluwer Acad. Publ. Manuf. Netherlands.*, Vol. 12343 Lncs, Pp. 5–32, 2001, Doi: 10.1007/978-3-030-62008-0_35.

[5] M. F. M. Khalik And F. Arifin, "Klasifikasi Indeks Kedalaman Kemiskinan Provinsi Sulawesi Selatan Berbasis Decision Tree, K-Nearest Neighbor, Naive Bayes, Neural Network, Dan Random Forest," *J. Edukasi Dan Penelit. Inform.*, Vol. 9, No. 2, P. 282, 2023, Doi: 10.26418/Jp.V9i2.67492.

[6] A. Heryati And T. Setiawan Saputra, "Optimizing Socioeconomic Features For Poverty Prediction In South Sumatera," *Tiers Inf. Technol. J.*, Vol. 6, No. 1, Pp. 16–32, 2025.

[7] N. Valentika, K. A. Notodiputro, And B. Sartono, "Performance Study Of Prediction Intervals With Random Forest For Poverty Data Analysis," *J. Apl. Stat. Komputasi Stat.*, Vol. 16, No. 1, Pp. 80–88, 2024, Doi: 10.34123/Jurnalasks.V16i1.542.

[8] S. Sobari, A. I. Purnamasari, A. Bahtiar, And K. Kaslani, "Meningkatkan Model Prediksi Kelulusan Santri Tahfidz Di Pondok Pesantren Al-Kautsar Menggunakan Algoritma Random Forest," *J. Inform. Dan Tek. Elektro Terap.*, Vol. 13, No. 1, 2025, Doi: 10.23960/Jitet.V13i1.5704.

[9] M. F. Y. Herjanto And Carudin, "Analisis Sentimen Ulasan Pengguna Aplikasi Sirekap Pada Play Store Menggunakan Algoritma Random Forest Classifer," Vol. 12, No. 2, Pp. 1204–1210, 2024.

[10] J. Foster, T. George, J. Greer, And S. Spring, "The Foster-Greer-Thorbecke Poverty Measures : Twenty Five Years Later By," Pp. 1–36.

[11] E. A. F. Elmuna, T. Chamidy, And F. Nugroho, "Optimization Of The Random Forest Method Using Principal Component Analysis To Predict House Prices," *Int. J. Adv. Data Inf. Syst.*, Vol. 4, No. 2, Pp. 155–166, 2023, Doi: 10.25008/Ijadis.V4i2.1290.

[12] D. N. Handayani And S. Qutub, "Penerapan Random Forest Untuk Prediksi Dan Analisis Kemiskinan," *Riggs J. Artif. Intell. Digit. Bus.*, Vol. 4, No. 2, Pp. 405–412, 2025, Doi: 10.31004/Riggs.V4i2.512.

[13] C. A. Mushagalusa, A. B. Fandohan, And R. Glèlè Kakaï, "Random Forests In Count Data Modelling: An Analysis Of The Influence Of Data Features And Overdispersion On Regression Performance," *J. Probab. Stat.*, Vol. 2022, Pp. 1–21, 2022, Doi: 10.1155/2022/2833537.

[14] S. Han, B. D. Williamson, And Y. Fong, "Improving Random Forest Predictions In Small Datasets From Two-Phase Sampling Designs," *Bmc Med. Inform. Decis. Mak.*, Vol. 21, No. 1, Pp. 1–9, 2021, Doi: 10.1186/S12911-021-01688-3.

[15] Vibhu Verma, "A Comprehensive Framework For Residual Analysis In Regression And Machine Learning," *J. Inf. Syst. Eng. Manag.*, Vol. 10, No. 31s, Pp. 34–46, 2025, Doi: 10.52783/Jisem.V10i31s.4958.