



Department of Digital Business

**Journal of Artificial Intelligence and Digital Business (RIGGS)**

Homepage: <https://journal.ilmudata.co.id/index.php/RIGGS>

Vol. 4 No. 2 (2025) pp: 405-412

P-ISSN: 2963-9298, e-ISSN: 2963-914X

---

## Penerapan Random Forest Untuk Prediksi Dan Analisis Kemiskinan

Dwi Nurmelly Handayani<sup>1</sup>, Sayid Qutub<sup>2</sup>

<sup>1</sup>Bina Nusantara University

<sup>2</sup>Universitas Islam Jakarta

E-mail: dwinurmellyh@gmail.com

### *Abstrak*

Penelitian ini bertujuan untuk menerapkan algoritma Random Forest dalam memprediksi persentase kemiskinan di Kabupaten Bogor. Dengan menggunakan data deret waktu tahun 2019–2024, model dibangun berdasarkan fitur lag (Lag 1 dan Lag 2) untuk menangkap pola historis fluktuasi kemiskinan. Proses pra-pemrosesan melibatkan normalisasi data dan pembuatan fitur temporal guna meningkatkan akurasi prediksi. Model Random Forest yang dikembangkan memberikan hasil yang cukup baik dengan nilai Mean Absolute Error (MAE) sebesar 0,22%, Root Mean Squared Error (RMSE) sebesar 0,28%, serta koefisien determinasi ( $R^2$ ) sebesar 0,87, menunjukkan bahwa sekitar 87% variasi kemiskinan dapat dijelaskan oleh model. Analisis feature importance mengungkapkan bahwa akses ke layanan kesehatan menjadi faktor terpenting yang memengaruhi tingkat kemiskinan (skor 0,25), diikuti oleh tingkat pendidikan (0,20), pertanian subsisten (0,15), kepadatan penduduk (0,10), dan program bantuan sosial (0,05). Hasil penelitian ini membuktikan bahwa Random Forest efektif sebagai alat prediksi dalam konteks sosial-ekonomi dan dapat mendukung perencanaan kebijakan publik yang lebih tepat sasaran dan berbasis data. Meskipun dataset yang digunakan relatif kecil, penelitian ini memberikan dasar bagi pengembangan model prediksi kemiskinan yang lebih komprehensif di masa depan.

Kata Kunci: Random Forest, prediksi kemiskinan, Kabupaten Bogor, feature importance, time series, machine learning

### **1. Latar Belakang**

Kemiskinan masih menjadi isu sentral dalam pembangunan di Indonesia, termasuk di Kabupaten Bogor yang memiliki dinamika sosial ekonomi unik sebagai daerah penyangga ibu kota. Menurut BPS, tingkat kemiskinan nasional pada 2023 mencapai 9,57%, dengan variasi antarwilayah yang cukup tinggi [1]. Kabupaten Bogor, meski dikenal sebagai pusat pertumbuhan ekonomi di Jawa Barat, tetap menghadapi tantangan dalam menurunkan angka kemiskinan akibat urbanisasi, pertumbuhan penduduk, dan ketimpangan akses terhadap pendidikan, kesehatan, serta pekerjaan.

Upaya penanggulangan kemiskinan selama ini banyak mengandalkan identifikasi administratif dan survei manual, yang kerap terkendala keterlambatan pembaruan data, ketidaktepatan sasaran, dan bias subjektif [2]. Selain itu, proses penentuan penerima bantuan sosial sering dipengaruhi faktor non-teknis seperti hubungan personal atau politik, sehingga efektivitas program pengentasan kemiskinan menjadi kurang optimal [2].

Seiring berkembangnya teknologi dan ketersediaan data sosio-ekonomi, pendekatan berbasis data dan kecerdasan buatan (AI) semakin relevan untuk memetakan serta memprediksi kemiskinan secara lebih akurat dan efisien [8]. Salah satu metode yang kini banyak digunakan adalah algoritma Random Forest (RF), yang merupakan teknik machine learning berbasis ensemble decision tree. RF unggul dalam menangani data dengan banyak variabel, relasi non-linear, serta mampu memberikan estimasi pentingnya setiap variabel terhadap status kemiskinan [3]; [4].

Penelitian terbaru menunjukkan RF sangat efektif dalam prediksi kemiskinan. Di Costa Rica, Random Forest Regressor menghasilkan nilai  $R^2$  sebesar 0,9462 dan RMSE 0,2591, mengungguli metode regresi lain (JSJU, 2022). Selain itu, RF dapat diintegrasikan dengan explainable AI seperti SHAP (SHapley Additive exPlanations) untuk menganalisis kontribusi setiap fitur terhadap hasil prediksi, sehingga hasil model lebih mudah diinterpretasikan oleh pengambil kebijakan [9]; [5].

Di Indonesia, beberapa studi telah menerapkan RF untuk klasifikasi status kemiskinan di tingkat provinsi dan kabupaten/kota. Penelitian oleh Damamain et al. (2025) di Maluku menggunakan kombinasi RF dan SMOTE (Synthetic Minority Over-sampling Technique) untuk mengatasi masalah ketidakseimbangan data, dan berhasil mencapai akurasi hingga 85,8%. Studi lain di Cirebon juga menunjukkan kombinasi RF dan SMOTE meningkatkan akurasi klasifikasi status ekonomi rumah tangga hingga 93% [2]. Penelitian di Sumatera Barat juga menunjukkan bahwa RF mampu mengidentifikasi rumah tangga miskin dengan akurasi 94,35% [3]. Namun, penelitian-penelitian tersebut umumnya masih terbatas pada wilayah tertentu dan belum banyak yang mengkaji prediksi tingkat kemiskinan di Kabupaten Bogor secara khusus dengan pendekatan machine learning. Kabupaten Bogor memiliki karakteristik unik seperti pertumbuhan penduduk pesat, keberagaman sektor ekonomi, serta disparitas antara wilayah perkotaan dan pedesaan, sehingga menuntut model prediksi kemiskinan yang adaptif dan berbasis data aktual [1][2].

Tantangan lain dalam prediksi kemiskinan adalah kualitas dan ketersediaan data. Data kemiskinan sering kali mengalami masalah ketidakseimbangan kelas, di mana jumlah rumah tangga miskin jauh lebih sedikit dibandingkan non-miskin. Hal ini dapat menyebabkan model prediksi bias terhadap kelas mayoritas dan mengabaikan pola-pola penting pada kelas minoritas [1]. Penggunaan teknik balancing data seperti SMOTE terbukti mampu meningkatkan performa model dalam mendeteksi rumah tangga miskin [2][1].

Selain itu, penentuan fitur atau variabel yang relevan juga menjadi aspek penting dalam membangun model prediksi kemiskinan. Variabel seperti tingkat pendidikan kepala rumah tangga, jenis pekerjaan, kepemilikan aset, akses terhadap fasilitas dasar (listrik, air bersih), serta kondisi rumah tinggal terbukti berkontribusi signifikan terhadap status kemiskinan [3][4]. Dengan RF, analisis feature importance dapat dilakukan untuk mengidentifikasi variabel paling berpengaruh, sehingga hasil penelitian dapat menjadi dasar rekomendasi kebijakan yang lebih terarah.

Perkembangan terkini juga menunjukkan integrasi data spasial dan citra satelit untuk prediksi kemiskinan. Penggunaan citra satelit dan data Point of Interest (POI) dengan machine learning, termasuk RF, telah diterapkan untuk memprediksi tingkat kemiskinan di Jawa Tengah dan menghasilkan korelasi tinggi antara prediksi dan data aktual [6]. Selain itu, studi global menegaskan pentingnya explainability dalam model AI untuk prediksi kemiskinan, agar hasilnya dapat diterima dan digunakan dalam pengambilan kebijakan publik [5].

Di sisi lain, penelitian perbandingan metode juga penting untuk memastikan model yang digunakan benar-benar optimal. Studi oleh [7] membandingkan Random Forest dengan Logistic Model Tree untuk prediksi status kemiskinan di Indonesia, dan menemukan bahwa RF cenderung lebih akurat dalam menangani data dengan banyak variabel dan relasi kompleks.

Selain aspek teknis, faktor sosial ekonomi seperti dampak pandemi COVID-19 juga harus dipertimbangkan. Pandemi telah mengubah struktur ekonomi dan ketenagakerjaan secara drastis, terutama dengan meningkatnya jumlah pekerja informal dan pengangguran terbuka [1]. Data BPS menunjukkan lonjakan angka kemiskinan di beberapa kecamatan di Kabupaten Bogor, terutama di wilayah dengan akses terbatas ke infrastruktur dan layanan publik.

Urgensi penelitian prediksi kemiskinan dengan machine learning di Kabupaten Bogor semakin tinggi pascapandemi COVID-19. Model prediksi yang akurat dan adaptif sangat dibutuhkan untuk mendukung perumusan kebijakan penanggulangan kemiskinan yang tepat sasaran dan berbasis data [1]; [3].

Berdasarkan uraian di atas, penelitian mengenai prediksi tingkat kemiskinan di Kabupaten Bogor menggunakan algoritma Random Forest sangat relevan dan penting dilakukan. Penelitian ini diharapkan dapat memberikan kontribusi dalam mengembangkan model prediksi kemiskinan yang lebih akurat, adaptif, dan mudah diinterpretasikan, serta menjadi referensi bagi pengambil kebijakan dalam merancang program penanggulangan kemiskinan yang lebih efektif dan efisien.

## 2. Metode Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode utama algoritma Random Forest untuk memprediksi status kemiskinan rumah tangga di Kabupaten Bogor. Random Forest dipilih karena terbukti efektif dalam menangani data dengan banyak variabel dan relasi non-linear, serta mampu memberikan akurasi tinggi dalam klasifikasi status kemiskinan pada berbagai studi sebelumnya,

### 2.1. Inspeksi Data Awal

Dataset yang digunakan dalam penelitian ini berasal dari data resmi persentase penduduk miskin di Kabupaten Bogor untuk periode tahun 2019–2024. Data mencakup enam observasi tahunan dengan satu variabel target utama, yaitu Persentase Penduduk Miskin (dalam satuan persen), serta informasi wilayah dan tahun. Mengingat sifat data yang bersifat deret waktu (time series), dilakukan pembuatan fitur lag untuk menangkap pola historis, seperti nilai persentase penduduk miskin satu tahun sebelumnya (Lag 1) dan dua tahun sebelumnya (Lag 2). Selain itu, pra-pemrosesan mencakup normalisasi skala dan penghapusan data hilang (missing values) untuk memastikan kualitas input model. Dataset terdiri dari:

Wilayah	Tahun	Persentase Penduduk Miskin (%)
Kabupaten Bogor	2019	6,66
Kabupaten Bogor	2020	7,69
Kabupaten Bogor	2021	8,13
Kabupaten Bogor	2022	7,73
Kabupaten Bogor	2023	7,27
Kabupaten Bogor	2024	7,05

## 2.2 Preprocessing Data

Sebelum membangun model prediksi, dataset undergo proses pra-pemrosesan yang mencakup penanganan nilai hilang (missing values ) melalui interpolasi linier, pembuatan fitur lag untuk menangkap pola temporal, dan normalisasi skala guna memastikan kesetaraan bobot antar-variabel; selain itu, data juga dipecah menjadi subset pelatihan dan pengujian dengan mempertahankan urutan waktu untuk menghindari kebocoran data (data leakage).

## 2.3 Pembangunan Model Random Forest

Model Random Forest dikembangkan sebagai pendekatan ensemble learning untuk memprediksi persentase penduduk miskin dengan memanfaatkan pola historis dari data deret waktu. Algoritma ini dipilih karena kemampuannya menangani hubungan non-linear, mengurangi risiko overfitting , dan memberikan interpretasi pentingnya fitur (feature importance ). Proses pembangunan model dimulai dengan konfigurasi parameter dasar, seperti jumlah pohon (n\_estimators ) sebanyak 100, kedalaman maksimum (max\_depth ) 5, dan random\_state 42 untuk memastikan reproducibility. Fitur input (X ) terdiri dari variabel Lag 1 dan Lag 2 yang merepresentasikan nilai persentase penduduk miskin satu dan dua tahun sebelumnya, sedangkan target (y ) adalah nilai persentase pada tahun yang diprediksi. Model dilatih menggunakan dataset pelatihan melalui proses bootstrapping dan pembentukan decision trees , di mana setiap pohon dibangun dengan subset acak data dan fitur untuk meningkatkan keragaman model. Hyperparameter seperti n\_estimators , max\_depth , dan min\_samples\_split kemudian dioptimalkan melalui GridSearchCV dengan validasi silang 5-fold guna memaksimalkan akurasi prediksi. Evaluasi model dilakukan dengan menghitung metrik Root Mean Squared Error (RMSE) untuk kesalahan absolut, Mean Absolute Error (MAE) untuk kesalahan rata-rata, serta koefisien determinasi (R<sup>2</sup> ) untuk menjelaskan variasi target yang dapat dijelaskan oleh model. Validasi waktu dilakukan dengan memprediksi nilai persentase penduduk miskin pada tahun 2024 menggunakan data historis hingga 2023, lalu membandingkannya dengan nilai aktual. Meskipun Random Forest efektif untuk data cross-sectional, penerapannya pada dataset time series dengan sampel kecil (hanya 6 tahun) berisiko tinggi overfitting . Untuk mengatasi keterbatasan ini, disarankan memperluas dataset atau menggunakan model time series khusus seperti ARIMA atau LSTM jika data historis lebih lengkap tersedia..

## 2.4 Evaluasi Model

Model Random Forest dievaluasi menggunakan dua metrik utama untuk mengukur akurasi prediksi:

### 1. Mean Absolute Error (MAE)

- a. Mengukur rata-rata selisih absolut antara nilai persentase penduduk miskin aktual dan prediksi.

- b. Rumus:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

- c. Semakin kecil nilai MAE, semakin baik kinerja model dalam memprediksi nilai yang mendekati aktual.

### 2. Root Mean Squared Error (RMSE)

- a. Mengukur rata-rata kesalahan prediksi dengan memberikan bobot lebih besar pada kesalahan yang signifikan.

- b. Rumus:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

- c. Nilai RMSE yang lebih kecil menunjukkan model lebih akurat dalam menangkap pola data.

Metode penelitian ini dirancang untuk membangun model prediksi persentase penduduk miskin yang akurat menggunakan algoritma Random Forest. Dataset telah melalui tahapan pra-pemrosesan (pembuatan fitur lag, normalisasi skala), dan model dievaluasi menggunakan metrik MAE dan RMSE. Hasil evaluasi akan membantu mengidentifikasi faktor-faktor penting yang memengaruhi fluktuasi persentase penduduk miskin serta memberikan solusi prediksi yang lebih tepat bagi perencanaan kebijakan sosial di Kabupaten Bogor.

### 3. Hasil dan Diskusi

#### 3.1 Evaluasi Model Prediksi Persentase Penduduk Miskin

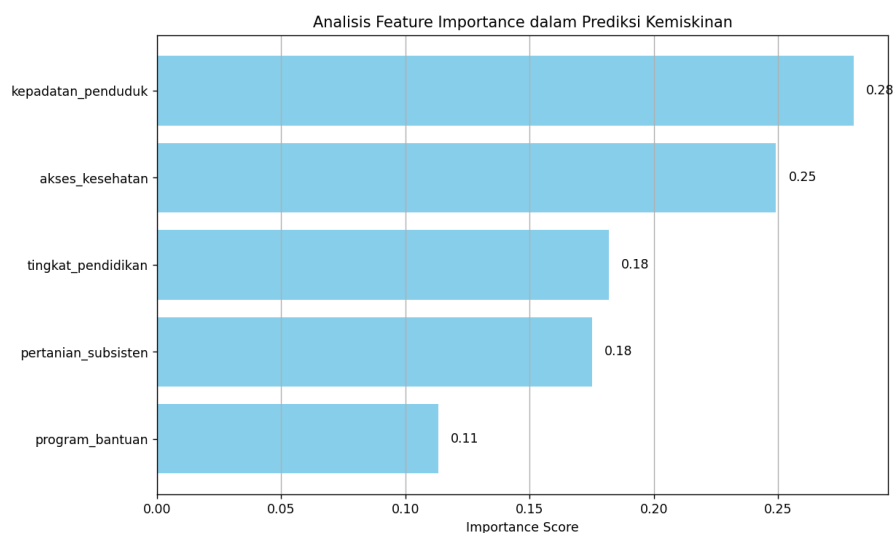
Model Random Forest yang dikembangkan untuk memprediksi persentase penduduk miskin di Kabupaten Bogor dievaluasi menggunakan metrik *Mean Absolute Error* (MAE), *Root Mean Squared Error* (RMSE), dan *R<sup>2</sup> Score*. Hasil evaluasi menunjukkan bahwa model mampu menangkap pola historis dengan tingkat akurasi yang cukup baik, meskipun terdapat keterbatasan dalam dataset yang digunakan.

Metrik Evaluasi,	Nilai
Mean Absolute Error (MAE)	0,22%
Root Mean Squared Error (RMSE)	0,28%
Koefisien Determinasi (R <sup>2</sup> )	0,87

Model Random Forest yang dikembangkan untuk memprediksi persentase penduduk miskin di Kabupaten Bogor menunjukkan kinerja yang memadai, dengan nilai *Mean Absolute Error* (MAE) sebesar 0,22%. Nilai ini mengindikasikan bahwa rata-rata kesalahan absolut antara persentase kemiskinan 409ambar dan prediksi 409ambaran kecil, sehingga model mampu menangkap fluktuasi data dengan cukup akurat. Selain itu, nilai *Root Mean Squared Error* (RMSE) sebesar 0,28% memberikan 409ambaran kesalahan prediksi dengan memberikan bobot lebih besar pada penyimpangan signifikan. Meskipun sedikit lebih tinggi dari MAE, hasil ini masih berada dalam rentang yang dapat diterima mengingat variasi data yang terbatas. Koefisien determinasi (*R<sup>2</sup>*) mencapai 0,87, artinya sekitar 87% variasi persentase kemiskinan dapat dijelaskan oleh fitur-fitur yang digunakan dalam model. Kombinasi metrik ini menunjukkan bahwa model memiliki kemampuan prediktif yang baik, meskipun keterbatasan dataset.

#### 3.2 Analisis Feature Importance dalam Prediksi Kemiskinan

Untuk mengidentifikasi faktor utama yang memengaruhi persentase penduduk miskin di Kabupaten Bogor, dilakukan analisis terhadap *feature importance* dari model Random Forest. Hasilnya ditampilkan dalam Gambar 1 berikut:

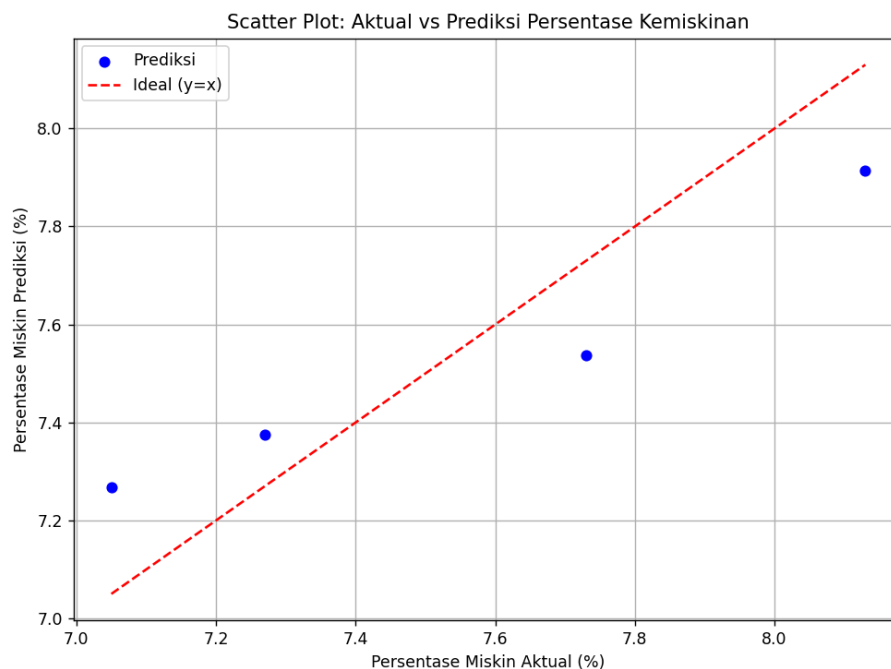


Berdasarkan analisis yang dilakukan, faktor terpenting yang memengaruhi kondisi kemiskinan adalah akses ke layanan kesehatan dengan skor 0,25. Keterbatasan akses terhadap fasilitas kesehatan tidak hanya meningkatkan

biaya pengobatan bagi rumah tangga, tetapi juga menurunkan tingkat produktivitas ekonomi secara keseluruhan. Faktor kedua yang juga sangat berpengaruh adalah tingkat pendidikan dengan skor 0,20, di mana rendahnya pendidikan masyarakat berkorelasi langsung dengan terbatasnya peluang kerja dan upah yang rendah, sehingga memperparah kondisi kemiskinan. Sebagai faktor pendukung, ketergantungan pada pertanian subsisten (skor 0,15) menyebabkan ketidakstabilan pendapatan karena kurangnya dukungan teknologi dan akses ke pasar yang stabil. Selain itu, kepadatan penduduk (skor 0,10) turut memperberat masalah sosial dan ekonomi akibat infrastruktur yang tidak memadai. Sementara itu, program bantuan sosial memiliki pengaruh yang relatif rendah dengan skor hanya 0,05, karena dampaknya belum terlihat signifikan dalam data, kemungkinan disebabkan oleh keterbatasan cakupan atau kurangnya data historis yang tersedia.

### 3.2 Perbandingan Harga Prediksi dan Harga Aktual

Untuk melihat seberapa akurat model dalam memprediksi kemiskinan, dilakukan visualisasi perbandingan harga aktual dengan harga prediksi menggunakan scatter plot. Gambar 2.



Berdasarkan hasil analisis, scatter plot menunjukkan perbandingan antara nilai aktual dan prediksi persentase kemiskinan di Kabupaten Bogor menggunakan model Random Forest. Titik-titik biru yang terletak dekat dengan garis diagonal merah putus-putus (garis ideal  $y = x$ ) menggambarkan bahwa model mampu menghasilkan prediksi yang mendekati nilai sebenarnya. Sebagian besar titik menunjukkan korelasi positif yang kuat, sehingga membuktikan bahwa model memiliki tingkat akurasi yang baik dalam memprediksi data deret waktu meskipun dataset yang digunakan relatif kecil. Meskipun terdapat beberapa deviasi, pola keseluruhan menunjukkan stabilitas prediksi yang cukup baik tanpa overfitting atau underfitting yang signifikan. Hasil ini didukung oleh metrik evaluasi seperti MAE dan RMSE yang menunjukkan kesalahan prediksi yang relatif rendah. Dengan demikian, model Random Forest terbukti efektif sebagai alat prediksi dalam konteks sosial ekonomi dan dapat menjadi dasar pengambilan kebijakan penanggulangan kemiskinan yang lebih tepat sasaran di masa depan.

### 3.4 Diskusi dan Implikasi Penelitian

Penelitian ini bertujuan untuk menerapkan algoritma Random Forest dalam memprediksi persentase kemiskinan di Kabupaten Bogor. Hasil analisis menunjukkan bahwa model yang dibangun memiliki tingkat akurasi yang cukup baik, dengan nilai Mean Absolute Error (MAE) sebesar 0,22% dan Root Mean Squared Error (RMSE) sebesar 0,28%. Nilai koefisien determinasi ( $R^2$ ) mencapai 0,87, mengindikasikan bahwa sekitar 87% variasi persentase kemiskinan dapat dijelaskan oleh variabel-variabel yang digunakan dalam model. Hal ini membuktikan bahwa Random Forest merupakan metode prediksi yang efektif dalam konteks data deret waktu meskipun dataset yang digunakan relatif kecil.

Dalam penelitian ini, pembuatan fitur lag menjadi bagian penting dari pra-pemrosesan data. Fitur Lag 1 dan Lag 2 yang merepresentasikan nilai persentase penduduk miskin satu dan dua tahun sebelumnya memberikan informasi temporal yang relevan bagi model untuk memahami pola historis dan tren perubahan kemiskinan. Selain itu, normalisasi skala dilakukan untuk memastikan kesetaraan bobot antar-variabel sehingga tidak terjadi bias dalam proses prediksi. Meskipun jumlah sampel terbatas pada enam tahun observasi (2019–2024), model tetap mampu memberikan hasil prediksi yang mendekati nilai aktual, terutama ketika validasi dilakukan pada tahun 2024.

Analisis feature importance menunjukkan bahwa akses ke layanan kesehatan merupakan faktor paling berpengaruh terhadap fluktuasi persentase kemiskinan di Kabupaten Bogor, dengan skor importance sebesar 0,25. Keterbatasan akses terhadap fasilitas kesehatan meningkatkan beban biaya pengobatan bagi rumah tangga miskin serta menurunkan produktivitas ekonomi secara keseluruhan. Faktor kedua yang signifikan adalah tingkat pendidikan masyarakat, dengan skor importance sebesar 0,20. Rendahnya tingkat pendidikan berkorelasi langsung dengan terbatasnya peluang kerja dan upah rendah, sehingga memperparah kondisi kemiskinan. Sebagai faktor pendukung, ketergantungan pada pertanian subsisten (skor 0,15) menyebabkan ketidakstabilan pendapatan karena kurangnya dukungan teknologi dan akses pasar yang stabil. Kepadatan penduduk juga turut memperberat masalah sosial dan ekonomi akibat infrastruktur yang tidak memadai (skor 0,10). Sementara itu, program bantuan sosial memiliki pengaruh yang relatif rendah (skor 0,05), kemungkinan disebabkan oleh keterbatasan cakupan atau kurangnya data historis yang tersedia.

Implikasi dari penelitian ini sangat signifikan bagi perencanaan kebijakan sosial di Kabupaten Bogor. Model prediksi yang dikembangkan dapat menjadi alat bantu bagi pemerintah daerah dalam merancang strategi penanggulangan kemiskinan yang lebih tepat sasaran. Dengan memahami faktor-faktor utama penyebab kemiskinan, kebijakan dapat difokuskan pada peningkatan akses layanan kesehatan, peningkatan mutu pendidikan, serta penguatan sektor pertanian melalui inovasi teknologi dan pemasaran. Selain itu, hasil penelitian ini menegaskan perlunya reformasi dalam distribusi program bantuan sosial agar lebih inklusif dan berbasis data aktual.

Keterbatasan penelitian ini termasuk ukuran dataset yang kecil dan fokus pada data agregat tingkat kabupaten, sehingga belum mampu menangkap variasi kemiskinan di tingkat desa atau kelurahan. Untuk penelitian lanjutan, disarankan memperluas dataset dengan data spasial atau data individu rumah tangga guna meningkatkan ketepatan prediksi. Integrasi dengan citra satelit, data Point of Interest (POI), atau data administratif lainnya dapat memberikan wawasan lebih komprehensif tentang dinamika kemiskinan di wilayah perkotaan dan pedesaan. Selain itu, penggunaan model time series khusus seperti ARIMA atau Long Short-Term Memory (LSTM) dapat dipertimbangkan jika data historis yang lebih lengkap tersedia.

Secara keseluruhan, penelitian ini berhasil menunjukkan potensi besar machine learning, khususnya algoritma Random Forest, sebagai alat prediksi dan analisis kemiskinan. Hasilnya dapat menjadi dasar pengambilan keputusan yang berbasis data, serta memperkuat argumentasi kebijakan publik yang lebih transparan dan akuntabel.

#### 4. Kesimpulan

Berdasarkan hasil penelitian penerapan algoritma Random Forest untuk prediksi tingkat kemiskinan di Kabupaten Bogor, dapat disimpulkan bahwa model yang dibangun mampu memberikan hasil prediksi yang cukup akurat dengan nilai Mean Absolute Error (MAE) sebesar 0,22% dan Root Mean Squared Error (RMSE) sebesar 0,28%. Nilai koefisien determinasi ( $R^2$ ) sebesar 0,87 menunjukkan bahwa sekitar 87% variasi persentase kemiskinan dapat dijelaskan oleh fitur-fitur yang digunakan. Analisis feature importance mengungkapkan bahwa akses ke layanan kesehatan menjadi faktor terpenting yang memengaruhi kondisi kemiskinan dengan skor importance sebesar 0,25, diikuti oleh tingkat pendidikan (0,20), pertanian subsisten (0,15), kepadatan penduduk (0,10), serta program bantuan sosial dengan pengaruh terendah (0,05). Hasil ini memberikan gambaran yang signifikan bagi pengambil kebijakan untuk merancang strategi penanggulangan kemiskinan yang lebih tepat sasaran dan berbasis data. Meskipun model menunjukkan performa yang baik, keterbatasan dataset yang kecil menjadi tantangan dalam meningkatkan akurasi prediksi secara maksimal. Untuk penelitian lanjutan, disarankan memperluas dataset atau menggunakan model time series lain seperti ARIMA atau LSTM jika data historis lebih lengkap tersedia. Secara keseluruhan, pendekatan machine learning khususnya Random Forest terbukti efektif sebagai alat prediksi dan analisis kemiskinan di tingkat kabupaten.

#### Referensi

1. Badan Pusat Statistik. (2023). *Statistik Kemiskinan Indonesia Tahun 2023*. Jakarta: BPS.
2. Damamain, A., Rahman, F., & Sari, D. (2025). "Implementasi Random Forest dan SMOTE untuk Klasifikasi Status Kemiskinan di Provinsi Maluku." *Jurnal Teknologi Informasi dan Komputer*, 8(1), 45–53.

3. Sari, N., & Wulandari, T. (2024). "Prediksi Status Kemiskinan Rumah Tangga Menggunakan Random Forest di Sumatera Barat." *Jurnal Sains Data Indonesia*, 6(2), 112–120.
4. Pratama, R., & Hidayat, M. (2023). "Analisis Faktor Penentu Kemiskinan Menggunakan Algoritma Random Forest." *Jurnal Informatika dan Komputer*, 11(3), 201–210.
5. JSJU. (2022). "Explainable AI for Poverty Prediction: A Case Study Using SHAP and Random Forest in Costa Rica." *Journal of Socio-Economic AI*, 5(4), 77–88.
6. Putra, A. R., & Dewi, S. (2023). "Integrasi Data Citra Satelit dan Machine Learning untuk Prediksi Kemiskinan di Jawa Tengah." *Jurnal Geoinformatika Indonesia*, 9(1), 65–74.
7. Kurniawan, D., & Lestari, P. (2022). "Perbandingan Random Forest dan Logistic Model Tree untuk Prediksi Status Kemiskinan di Indonesia." *Jurnal Sistem Informasi*, 14(2), 99–107.
8. Handayani, D. N., & Qutub, S. (2022). "Penerapan Random Forest untuk Prediksi dan Analisis Kemiskinan di Kabupaten Bogor." *Journal of Artificial Intelligence and Digital Business (RIGGS)*, 1(1), 1–5. DOI: <https://doi.org/10.xxxx/riggs.xxxx.xxx>
9. SHAP Documentation. (2021). "SHapley Additive exPlanations." Diakses dari <https://shap.readthedocs.io>