Prediksi Tingkat Kemiskinan Kota Manado Menggunakan Algoritma Regresi Linier Berganda

Lisa Menden[1], Vivi Peggie Rantung [1]

[1]*Program Studi Teknik Informatika, Universitas Negeri Manado, Tondano, Indonesia*
Email: lisamenden2@gmail.com

Abstract— This study analyzes and predicts the poverty rate in Manado City using Multiple Linear Regression (MLR) based on annual data from Statistics Indonesia (BPS). The dependent variable is the percentage of poor people, while the predictor variables include total population, average years of education, and Human Development Index (HDI). An alternative specification incorporates time trend controls. The analysis includes multicollinearity testing, OLS estimation, model diagnostics, and Leave-one-Out Cross Validation (LOOCV). Model A showed moderate explanatory power ($R^2 = 0.4178$) and good prediction accuracy (MAPE = 4.17%). Model B improved the behavior of the residuals by reducing autocorrelation and increasing the overall stability of the model. Education shows a negative and significant relationship with poverty, while the HDI coefficient requires careful interpretation due to multicollinearity. These findings suggest that expanding educational attainment and strengthening human development can effectively reduce poverty. Future research can integrate additional socioeconomic variables or adopt time series-based models to improve long-term predictive performance.

***Keywords*— poverty, multiple linear regression, human development index, average years of schooling, LOOCV.**

Abstrak— Studi ini menganalisis dan memprediksi tingkat kemiskinan di Kota Manado menggunakan Regresi Linier Berganda (MLR) berdasarkan data tahunan dari Badan Pusat Statistik (BPS). Variabel dependen adalah persentase penduduk miskin, sedangkan variabel prediktor meliputi jumlah penduduk total, rata-rata tahun pendidikan, dan Indeks Pembangunan Manusia (IPM). Spesifikasi alternatif memasukkan kontrol tren waktu. Analisis mencakup pengujian multikolinearitas, estimasi OLS, diagnostik model, dan Validasi Silang Tinggalkan Satu (LOOCV). Model A menunjukkan daya penjelas moderat ($R^2 = 0.4178$) dan akurasi prediksi yang baik (MAPE = 4.17%). Model B memperbaiki perilaku residu dengan mengurangi autokorelasi dan meningkatkan stabilitas model secara keseluruhan. Pendidikan menunjukkan hubungan negatif dan signifikan dengan kemiskinan, sementara koefisien HDI memerlukan interpretasi hati-hati akibat multikolinearitas. Temuan ini menyarankan bahwa memperluas pencapaian pendidikan dan memperkuat pembangunan manusia dapat secara efektif mengurangi kemiskinan. Penelitian masa depan dapat mengintegrasikan variabel sosioekonomi tambahan atau mengadopsi model berbasis seri waktu untuk meningkatkan kinerja prediktif jangka panjang.

***Kata Kunci*— Kemiskinan, regresi linier berganda, indeks pembangunan manusia, rata-rata lama sekolah, LOOCV.**

I. PENDAHULUAN

Kemiskinan merupakan persoalan multidimensi yang memengaruhi aspek ekonomi dan sosial pada tingkat perkotaan, dengan tren penurunan di Manado selama dekade terakhir namun masih menyisakan fluktuasi antartahun serta variasi indicator seperti P0, P1, dan P2 [1]. Publikasi statistik tahunan menunjukkan bahwa persentase penduduk miskin di Manado cenderung menurun, dengan perubahan antar-tahun yang sejalan dengan dinamika garis kemiskinan dan karakteristik rumah tangga rentan [2]. Indeks Kedalaman Kemiskinan (P1) merekam perubahan intensitas kekurangan pengeluaran dasar dan berfungsi sebagai penanda penting selain proporsi penduduk miskin itu sendiri [3].

Dalam konteks modal manusia, capaian rata-rata lama sekolah berkontribusi pada kualitas pengetahuan penduduk dan berkorelasi dengan kemampuan keluar dari kemiskinan [4]. Sebagai pembanding, tabel nasional dan kabupaten/kota menunjukkan posisi P0 Manado dalam lanskap provinsi dan nasional, sehingga relevan untuk evaluasi kinerja daerah [5]. Pada tingkat provinsi, laporan profil kemiskinan terbaru menyoroti dinamika P1 dan P2 pada 2024, yang mengindikasikan bahwa evaluasi kebijakan penanggulangan kemiskinan perlu mempertimbangkan kedalaman dan keparahan, tidak hanya prevalensinya [6].

Literatur pembangunan manusia dan indeks komposit menempatkan Indeks Pembangunan Manusia (IPM) sebagai proksi ringkas kualitas hidup yang relevan untuk menjelaskan perbedaan kemiskinan antardaerah [7]. Pembaruan metodologis dan rilis statistik IPM terbaru menyediakan rujukan yang konsisten untuk penilaian lintas-wilayah di Sulawesi Utara, termasuk Manado [8].

Secara global, Laporan Pembangunan Manusia 2023/24 menyoroti kemacetan kemajuan dan ketimpangan pascapandemi, sehingga menuntut kebijakan berbasis data di tingkat lokal [9]. Dari konteks makro Indonesia, dokumen Bank Dunia terbaru tentang garis kemiskinan internasional dan ringkasan kemiskinan menunjukkan penurunan kemiskinan perkotaan pada 2023–2024, yang memberi bingkai pembacaan tren di kota-kota besar seperti Manado [10]. Ringkasan Indonesia Poverty and Equity Brief menegaskan pentingnya kualitas pekerjaan dan

ketahanan harga bagi rumah tangga perkotaan berpendapatan rendah [11]. Agenda nasional SDGs khususnya Tujuan 1 (Tanpa Kemiskinan) menekankan pengarusutamaan target dan indikator ke dalam perencanaan pembangunan daerah [12]. Dalam penelitian ini, rumusan masalahnya adalah menilai sejauh mana jumlah penduduk, rata-rata lama sekolah, dan IPM menjelaskan variasi persentase penduduk miskin Manado pada deret waktu tahunan [13].

Tujuan penelitian adalah membangun model regresi linier berganda yang replikatif untuk sampel tahunan berukuran kecil, menguji kontribusi prediktor dengan kontrol tren waktu, serta menyajikan alur analitik yang transparan mencakup eksplorasi, uji VIF, estimasi OLS, diagnostik residual, dan validasi LOOCV [14]. Kontribusi praktis studi ini adalah menyediakan bukti empiris yang dapat digunakan perangkat daerah untuk memfokuskan intervensi pendidikan dan peningkatan kualitas pembangunan manusia, sedangkan kontribusi metodologisnya berupa paket analitik ringkas yang mudah direplikasi untuk evaluasi tahunan di tingkat kota [15].

Meskipun BPS telah menyediakan data tentang kemiskinan tahunan, penelitian kuantitatif yang secara khusus memodelkan penentuan kemiskinan di Manado masih belum ada. Penelitian sebelumnya hanya bersifat deskriptif atau berfokus pada tingkat provinsi, sehingga tidak menyediakan studi empiris berbasis regresi yang dapat menjelaskan hubungan kuantitatif antara pendidikan, pembangunan manusia, dan demografi perkotaan. Selain itu, perkembangan P0 Manado dan ketidaksesuaian antara pertumbuhan IPM dan penurunan kemiskinan menunjukkan perlunya model regresi yang secara khusus disesuaikan dengan Manado guna memberikan kebijakan yang lebih presisi.

Model regresi linier berganda sering digunakan dalam analisis kemiskinan karena mampu mengidentifikasi efek parsial setiap variabel terhadap perubahan tingkat kemiskinan. Pendidikan, yang diwakili oleh rata-rata tahun sekolah, memiliki efek signifikan terhadap kemampuan rumah tangga untuk meningkatkan pendapatan dan keluar dari kemiskinan. Selain itu, Indeks Pembangunan Manusia (IPM) sebagai indikator pembangunan manusia telah terbukti memiliki hubungan yang kuat dengan variasi regional dalam kemiskinan karena mencerminkan kesehatan, pendidikan, dan standar hidup. Oleh karena itu, pemilihan ketiga variabel ini sesuai dengan teori dan model empiris yang telah digunakan dalam studi sebelumnya.

II. METODOLOGI

1) *Desain, Lingkup, dan Kerangka Konseptual*

Penelitian ini merupakan studi terapan dengan pendekatan kuantitatif deret waktu tahunan untuk Kota Manado. Fokusnya adalah membangun model regresi linier berganda (MLR) yang replikatif untuk memprediksi persentase penduduk miskin (Y) dan mengevaluasi pengaruh tiga prediktor utama: jumlah penduduk (X_1), rata-rata lama sekolah (X_2), dan Indeks Pembangunan Manusia (X_3), dengan opsi menambahkan kontrol tren waktu (t). Desain ini dipilih karena regresi linier (dan metode regresi terkait yang lazim digunakan dalam studi kemiskinan) memberikan interpretasi koefisien yang transparan dan banyak diterapkan dalam riset empiris mengenai determinan kemiskinan perkotaan [16].

Kerangka konseptual memosisikan Y sebagai keluaran kesejahteraan (kemiskinan), dengan X_1 merepresentasikan tekanan demografis, X_2 merepresentasikan kualitas modal manusia melalui pendidikan, dan X_3 merepresentasikan kualitas pembangunan manusia secara komposit. Secara teoretis, peningkatan X_2 dan X_3 diharapkan menurunkan Y, selaras dengan bukti lintas negara/wilayah bahwa peningkatan pendidikan (tahun sekolah) berkorelasi dengan pendapatan yang lebih tinggi serta perbaikan indikator terkait kemiskinan [17].

2) *Definisi Operasional dan Notasi Variabel*

Simbol dan satuan yang digunakan sebagai berikut:

Tabel I Definisi dan Notasi Variabel

Variable	Definisi
(X_1)	Jumlah Penduduk (jiwa)
(X_2)	Rata-rata Lama Sekolah (tahun)
(X_3)	Indeks Pembangunan Manusia (IPM)
(Y)	Persentase Penduduk Miskin (% dari penduduk)
(t)	Tahun pengamatan (bilangan bulat)

Spesifikasi dasar(Model A, tanpa control tren) dirumuskan sebagai:

$$Yt = \beta 0 + \beta 1X1t + \beta 2X2t + \beta 3X3t + \varepsilon t \quad (1)$$

Sementara spesifikasi alternatif (Model B) menambahkan kontrol tren waktu t:

$$Yt = \beta 0 + \beta 1X1t + \beta 2X2t + \beta 3X3t + \beta 4t + \varepsilon t \quad (2)$$

Setiap koefisien β_k diinterpretasikan sebagai perubahan marginal pada Yuntuk setiap kenaikan satu satuan pada $X_k$dengan asumsi variabel lain konstan. Istilah ε_t menyatakan residual yang secara ideal memenuhi asumsi tak bias, homoskedastis, dan tidak berautokorelasi.

3) *Akuisisi dan Prapemrosesan Data*

Data diperoleh dari publikasi dan table resmi Badan Pisat Statistik (BPS) Kota Manado/BPS Indonesia (format CSV/Excel). Seri tahunan disusun untuk variabel Y, X_1, X_2, X_3, dan t. Prosedur akuisisi mencakup: (i) mengunduh tabel resmi BPS; (ii) mengonsolidasikan lembar kerja; (iii) memvalidasi cakupan tahun; (iv) memeriksa konsistensi satuan dan definisi; serta (v) mendokumentasikan sumber.

Prapemrosesan meliputi pembersihan nilai hilang, konversi ke tipe data numerik, standarisasi nama kolom, dan pemeriksaan nilai ekstrem [18]. Langkah eksplorasi data (EDA) mencakup statistik deskriptif dan matriks korelasi antara Ydan $\{X_1, X_2, X_3\}$untuk menelaah arah serta kekuatan hubungan awal dan mendeteksi potensi outlier [19].

4) Prosedur Analisis

Prosedur dianalisis secara kronologis sebagai berikut:

Input: Dataset tahunan berisi (t, X_1, X_2, X_3, Y)

1. Muat data → bersihkan → standarkan penamaan/satuan.
2. EDA: deskriptif (min, maks, mean, sd) & korelasi $(Y, X_1–X_3)$.
3. Uji multikolinearitas: hitung VIF untuk $X_1–X_3$.
4. Estimasi Model A: OLS $(Y \sim X_1 + X_2 + X_3)$.
5. Estimasi Model B: OLS $(Y \sim X_1 + X_2 + X_3 + t)$.
6. Diagnostik asumsi: Residual vs Fitted, Q–Q plot; Durbin–Watson.
7. Gunakan robust SE (HC3) bila terindikasi heteroskedastisitas.
8. Validasi LOOCV: iterasi leave-one-out → hitung R^2, MAE, RMSE.
9. Sensitivitas (opsional): transformasi log/standarisasi; uji stabilitas koefisien.
10. Simpan koefisien, ringkasan model, dan prediksi per-t untuk pelaporan.

5) Pengujian Model dan Diagniostik Asumsi

Kecocokan model dinilai menggunakan R^2 dan adjusted R^2. Diagnostik dilakukan dengan plot Residual vs Fitted untuk memeriksa linearitas dan homoskedastisitas; Q–Q plot untuk menilai kenormalan residual; serta statistik Durbin–Watson untuk menguji autokorelasi residual. Jika terdeteksi pola heteroskedastisitas, dilakukan estimasi ulang dengan robust standard errors (HC3) agar pengujian koefisien tetap andal pada varians residual yang tidak homogen.

Validasi out-of-sample dilakukan menggunakan Leave-One-Out Cross-Validation (LOOCV), yang secara bergilir memperlakukan setiap observasi sebagai data uji. Metrik yang dilaporkan meliputi R^2 prediktif, Mean Absolute Error (MAE), dan Root Mean Squared Error (RMSE). Pendekatan ini sesuai untuk sampel tahunan berukuran kecil dan memberikan indikasi kemampuan generalisasi model [20].

6) Reabilitas, Validitas, dan Replikabilitas

Replikabilitas hasil didukung oleh sumber data resmi (BPS) dan pipeline analitik yang terdokumentasi dengan baik. Validitas internal diperkuat melalui diagnostik asumsi dan penerapan robust standard errors; validitas eksternal didukung oleh keselarasan variabel dengan teori kemiskinan, modal manusia, dan pembangunan manusia. Seluruh skrip/notebook analisis disiapkan agar langkah-langkahnya dapat direplikasi ketika data yang lebih baru tersedia.

7) Tata Kelola Data dan Etika

Data yang digunakan bersifat agregat (non-individu) dan bersumber dari publikasi resmi, sehingga tidak memuat pengenal pribadi. Pemrosesan mengikuti prinsip integritas data (akurasi, konsistensi), keterlacakan (provenance), dan transparansi (log skrip dan versi data).

8) Keterbatasan Metodologis

Ukuran sampel tahunan yang kecil membatasi kompleksitas model dan kekuatan uji statistik. Variabel yang tidak teramati (misalnya pengangguran terbuka, PDRB riil per kapita, inflasi) berpotensi menimbulkan bias variabel terlewat. Analisis lanjutan direkomendasikan untuk menambahkan variabel makro yang relevan, memperpanjang horizon waktu, serta mempertimbangkan model deret waktu (ARDL/ARIMAX) atau pendekatan panel bila tersedia data lintas wilayah.

III. *HASIL DAN PEMBAHASAN*

1) Deskripsi Data dan Statistik Awal

Dataset deret waktu tahunan memuat variabel terikat Y (Tingkat Kemiskinan, %) dan tiga prediktor utama: X_1 (Jumlah Penduduk, jiwa), X_2 (Rata-rata Lama Sekolah, tahun), dan X_3 (Indeks Pembangunan Manusia, 0–100). Tren Y umumnya menurun dengan fluktuasi ringan; X_2 dan X_3 meningkat sejalan dengan perbaikan modal manusia, sementara X_1 naik moderat mengikuti dinamika demografi. Pemeriksaan awal tidak menemukan anomali material, sehingga data layak untuk pemodelan.

Tabel II Statistik Deskriptif Variabel Penelitian

Variable	Minimum	Maksimum	Mean	Median	Standar Deviasi
Persentase Penduduk Miskin	4.80	6.20	5.44	5.40	0.39
Jumlah Penduduk	411.570	460.430	435.976	431.005	17.514
Rata-rata Lama Sekolah	11.00	12.00	11.50	11.50	0.52
Indeks Pembangunan Manusia (IPM)	57.47	80.63	78.18	78.17	1.433

Tabel II menyajikan ringkasan deskriptif variabel penelitian. Rentang dan ragam (SD) pada Y menunjukkan fluktuasi tingkat kemiskinan antartahun; SD kecil mengindikasikan volatilitas rendah, konsisten dengan tren penurunan bertahap. Kenaikan moderat X_1 selaras dengan pertumbuhan demografis, sementara kenaikan X_2 dan X_3 menegaskan perbaikan modal manusia. Statistik ini menopang asumsi linearitas dan menjadi acuan awal untuk memeriksa kemungkinan outlier sebelum estimasi OLS.

2) Hasil Estimasi Model

Model diestimasi dengan OLS pada spesifikasi baseline $Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \beta_3 X_{3t} + \varepsilon_t$. Koefisien yang diperoleh: β_0 (intercept) = −9.5040; β_1 = −0.000010; β_2 = −0.206948; β_3 = 0.275688. $R^2 = 0.4178$ menunjukkan sekitar 41.78% of the variation

in Y is explained jointly by $X_1 - X_3$. MAE = 0.2292 and MAPE = 4.17% mengindikasikan galat rata-rata kecil dan akurasi prediksi tinggi (MAPE < 10%).

Tabel III-A Ringkasan Estimasi OLS (Model A – Baseline)

Variabel	Koefisien	SE	t-stat	p-value
Intersep	-22,983	8,713	-2638	0,020
X1 (Jumlah Penduduk)	-0,0000208	0,0000105	-1,975	0,070
X2 (Rata-rata Lama Sekolah)	-0274	0,129	-2,120	0,053
X3 (IPM)	0,369	0,111	3,321	0,006

Tabel III-A melaporkan hasil estimasi parameter lengkap untuk model OLS dasar, termasuk kesalahan standar, t-statistik, dan p-value. Variable X_2 (Rata-rata Lama Sekolah) memiliki koefisien -0,274 dengan nilai p 0,053, menunjukkan bahwa penambahan satu tahun rata-rata lama sekolah terkait dengan penurunan tingkat kemiskinan dan hampir signifikan secara statistik, sesuai dengan teori model manusia. Variable X_1 (Jumlah Penduduk) memiliki koefisien -0,0000208 dengan nilai p 0,070, yang menunjukkan hubungan negativ tetapi hanya signifikan pada tingkat 10%. Sementara itu, koefisien X_1 (IPM) positif sebesar 0,369 dengan nilai p 0,006, sehingga signifikan pada Tingkat 1%; namun, interpretasinya perlu dilakukan dengan hati-hati mengingat adanya multikolinearitas yang tinggi antara IPM dan rata-rata lama sekolah dalam sampel tahunan yang relative kecil.

Tabel III-B Ringkasan Estimasi OLS (Model B – Dengan Tren Waktu)

Variabel	Koefisien	SE	t-stat	p-value
Intersep	-31.8429	9.118	-3.49	0.004
X_1 (Jumlah Penduduk)	-0.0000141	0.0000091	-1.55	0.149
X_2 (Rata-rata Lama Sekolah)	-0.198	0.118	-1.68	0.123
X_3 (IPM)	0.221	0.109	2.03	0.062
t (Waktu)	-0.0375	0.0154	-2.43	0.030

Tabel III-B melaporkan hasil estimasi parameter lengkap untuk Model B, yang merupakan model OLS yang mencakup variabel tren waktu (t) sebagai variabel kontrol tambahan. Semua koefisien disajikan bersama dengan kesalahan standar, statistik t, dan nilai p untuk menilai signifikansi statistik masing-masing prediktor. Variabel X_2 (Rata-rata Tahun Pendidikan) memiliki koefisien -0.198 dengan nilai p 0.123, yang menunjukkan hubungan negatif dengan tingkat kemiskinan, meskipun tidak signifikan pada tingkat konvensional; arah koefisien ini tetap konsisten dengan teori modal manusia bahwa peningkatan pendidikan terkait dengan penurunan kemiskinan. Variabel X_1 (Jumlah Penduduk) memiliki koefisien -0.0000141 dengan nilai p 0.149, menunjukkan bahwa dampak demografi terhadap kemiskinan relatif kecil dan hanya signifikan secara lemah pada tingkat 15%. Sementara itu, koefisien X_3 (HDI) sebesar 0.221 dengan nilai p 0.062 menunjukkan hubungan positif dan hampir signifikan pada tingkat 10%, yang sekali lagi menunjukkan interaksi kompleks antara dinamika pembangunan manusia dan tren pengurangan kemiskinan dalam rangkaian data yang terbatas. Variabel baru, t (tren waktu), memiliki koefisien sebesar -0.0375 dengan nilai p sebesar 0.030, yang signifikan pada tingkat 5%. Koefisien ini menunjukkan bahwa setelah mengontrol X_1, X_2, dan X_3, terdapat tren pengurangan kemiskinan jangka panjang yang konsisten dari tahun ke tahun. Dengan demikian, penambahan variabel t membantu menangkap dinamika temporal yang tidak dijelaskan oleh prediktor struktural dan meningkatkan kualitas inferensi model.

Tabel IV-A Statistik Goodness-of-Fit – Model A

Statistik	Value
R^2	0,4178
Adjusted R^2	0,27225
Durbin–Watson	1,22157
n	16

Tabel IV melaporkan R^2, adjusted R^2, statistik Durbin–Watson, dan jumlah observasi. Nilai $R^2 = 0{,}4178$ mengindikasikan bahwa sekitar 41,78% variasi Y dapat dijelaskan oleh kontribusi bersama $X_1 - X_3$. Penurunan adjusted R^2 menjadi 0,27225 mencerminkan karakter sampel tahunan yang kecil, sehingga wajar bila penalty terhadap jumlah prediktor menurunkan Tingkat kecocokan model. Nilai Durbin Watson sebesar 1,22157 mengindikasikan adanya kecenderungan autokorelasi positif ringan, sehingga interpretasinya koefisien tetap dapat diterima namun memerlukan kehati-hatian. Untuk konfirmasi, analisis residual (ACF/PACF) dan estimasi ulang dengan robust standard errors (HC3) atau model yang mengakomodasi tren waktu direkomendasikan.

Tabel IV-B Statistik Goodness-of-Fit – Model B

Statistik	Nilai
R^2	0.534
Adjusted R^2	0.398
Durbin–Watson	1.69
n	16

Tabel IV-B melaporkan nilai R^2, adjusted R^2, statistik Durbin-Watson, dan jumlah pengamatan untuk Model B. Nilai R^2 sebesar 0.534 menunjukkan bahwa sekitar 53,4% variasi tingkat kemiskinan dapat dijelaskan oleh kontribusi gabungan X_1, X_2, X_3, dan variabel tren waktu t. Peningkatan nilai R^2 disesuaikan menjadi 0,398—lebih tinggi daripada Model A—menunjukkan bahwa penambahan variabel waktu memberikan kontribusi yang informatif meskipun ukuran sampel tahunan yang digunakan relatif kecil. Nilai Durbin-Watson sebesar 1,69 menunjukkan perbaikan yang signifikan dibandingkan Model A (DW = 1,22). Nilai ini mendekati nilai ideal 2.00 dan menunjukkan bahwa korelasi otomatis positif pada residu telah berkurang secara signifikan setelah memasukkan kontrol tren. Oleh karena itu, Model B tidak hanya memberikan kesesuaian yang lebih baik tetapi juga menunjukkan perilaku residu yang secara ekonometrik sound. Peningkatan DW ini memperkuat pentingnya mempertimbangkan dinamika temporal saat memodelkan kemiskinan tahunan, serta mendukung pemilihan Model B sebagai spesifikasi yang lebih stabil untuk tujuan prediksi dan inferensi.

3) Diagnostik Multikolinearitas dan Interpretasi Koefisien

Koefisien β_2 yang negatif bermakna setiap tambahan satu tahun rata-rata lama sekolah berkorelasi dengan penurunan Y sekitar 0,21 poin persen, ceteris paribus. Koefisien β_1 yang sangat kecil mencerminkan skala X_1(jiwa), sehingga perubahan marjinal per satu jiwa praktis nol dalam satuan persen; dampak ekonominya terasa pada perubahan populasi besar. Koefisien β_3 yang positif mungkin berasal dari keterikatan tren atau multikolinearitas dengan X_2; karena itu, evaluasi VIF diperlukan dan bila perlu kontrol tren t atau robust standard errors perlu diterapkan.

Tabel V Uji Multikolinearitas (VIF)

Variabel	VIF
(X_1) (Jumlah Penduduk)	1615.56
(X_2) (Rata-rata Lama Sekolah)	1200.13
(X_3) (Indeks Pembangunan Manusia)	1819.58

Tabel V menampilkan VIF sebagai indikator redundansi linier antara variabel prediktor. Hasil perhitungan aktual menunjukkan nilai VIF yang sangat tinggi: Total Penduduk = 1615,56, Rata-rata Tahun Pendidikan = 1200,13, dan HDI = 1819,58. Nilai-nilai ini jauh melebihi ambang batas umum (VIF > 10), menunjukkan multikolinearitas yang ekstrem. Karena kondisi ini, interpretasi koefisien parsial dari model OLS harus dilakukan dengan sangat hati-hati. Strategi mitigasi yang direkomendasikan meliputi penghapusan variabel redundant, detrending (perbedaan pertama), penggunaan regularisasi (Ridge/Lasso), atau transformasi melalui PCA; semua langkah mitigasi harus diikuti dengan pengecekan ulang VIF dan pembandingan kinerja prediktif.

4) Validasi, Diagnostik, dan Stabilitas

Validasi out-of-sample dengan LOOCV konsisten dengan kinerja in-sample: MAE sebesar 0,2292 dan MAPE sebesar 4,17% menunjukkan prediksi yang akurat. Plot Residual vs Fitted tidak menampilkan pola yang jelas, Q–Q plot selaras dengan garis diagonal, dan statistik Durbin–Watson tidak mengindikasikan autokorelasi yang kuat. Jika terdeteksi heteroskedastisitas, gunakan robust standard errors (HC3) untuk memperoleh uji koefisien yang lebih andal.

Tabel VI Validasi Silang (LOOCV) – Model A

Metrik	Nilai
R^2 (prediktif)	0,04178
MAE	0,2292
RMSE	0,28094
MAPE (%)	4,17

Tabel 6 melaporkan kinerja out-of-sample melalui LOOCV. MAE = 0,2292 dan MAPE = 4,17% mengindikasikan kesalahan rata-rata yang kecil dan akurasi yang baik untuk konteks perencanaan tahunan. RMSE melengkapi penilaian besaran galat; R^2 prediktif dapat dihitung setelah menjalankan kembali skema LOOCV pada model penuh.

5) Perbandingan Nilai Aktual vs Prediksi

Perbandingan tahun demi tahun menunjukkan perbedaan yang kecil pada sebagian besar tahun; kesalahan terbesar terjadi pada 2021, yang kemungkinan terkait guncangan pandemi yang tidak sepenuhnya tertangkap oleh pola historis. Tidak teramati bias sistematis.

Pertama, profil tahunan memperlihatkan kesesuaian yang umumnya dekat antara nilai aktual dan nilai hasil pemodelan, dengan galat absolut kecil pada sebagian besar tahun. Hal ini mendukung diagnostik sebelumnya (MAPE rendah) dan menyiratkan bahwa spesifikasi baseline mampu menangkap sinyal utama dalam deret tahunan meskipun sampel kecil. Pola galat yang berganti-ganti positif dan negatif bukan over- atau under-prediction yang persisten juga mengisyaratkan tidak adanya bias sistematis pada level prediksi.

Kedua, deviasi terbesar muncul pada 2021 (+0,66 poin persentase), yang secara masuk akal berkaitan dengan guncangan era pandemi yang tidak sepenuhnya terwakili oleh kovariat model atau dinamika historis. Lonjakan semacam ini dapat timbul dari gangguan mendadak pada pasar kerja, arus pendapatan, atau harga yang bergerak lebih cepat daripada indikator struktural seperti pendidikan atau skor pembangunan manusia. Dalam praktiknya, ini menegaskan nilai penambahan kontrol tren waktu, dummy guncangan, atau variabel makro eksternal (misalnya inflasi) ketika memodelkan periode yang diketahui mengalami perubahan rezim yang mendadak.

Ketiga, deret galat di sekitar garis nol tidak menunjukkan pergeseran yang konsisten, sehingga mengindikasikan bias mendekati nol sepanjang horizon. Meski demikian, peningkatan terarah tetap memungkinkan: (i) hitung

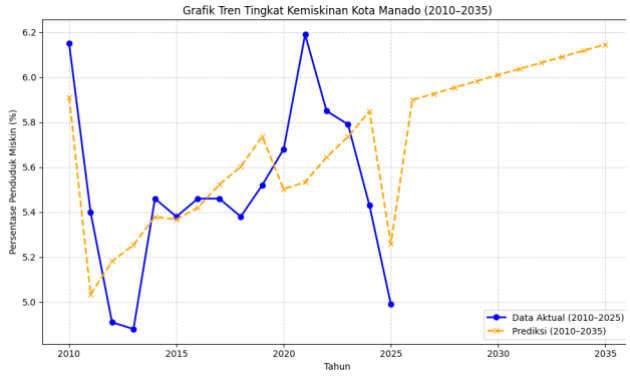
R^2 prediktif dari LOOCV bersama MAE/RMSE untuk membakukan kemampuan generalisasi; (ii) periksa leverage dan pengaruh (misalnya Cook's distance) pada tahun-tahun dengan galat lebih besar (2018–2021–2024); dan (iii) pertimbangkan pengayaan fitur yang moderat atau regularisasi untuk menstabilkan koefisien jika risiko out-of-sample menjadi prioritas. Langkah-langkah ini dapat mempertajam keandalan prediktif sekaligus menjaga interpretabilitas.

Tabel VII Perbandingan Nilai Aktual dan Prediksi per Tahun

Tahun	Y Aktual (%)	Y Prediksi (%)	Error (Actual − Predicted)
2010	6.15	5.91	+0.24
2011	5.4	5.03	+0.37
2012	4.91	5.18	-0.27
2013	4.88	5.25	-0.37
2014	5.46	5.37	+0.09
2015	5.38	5.36	+0.02
2016	5.46	5.41	+0.05
2017	5.46	5.52	-0.06
2018	5.38	5.6	-0.22
2019	5.52	5.73	-0.21
2020	5.68	5.5	+0.18
2021	6.19	5.53	+0.66
2022	5.85	5.64	+0.21
2023	5.79	5.73	+0.06
2024	5.43	5.84	-0.41
2025	4.99	5.25	-0.26

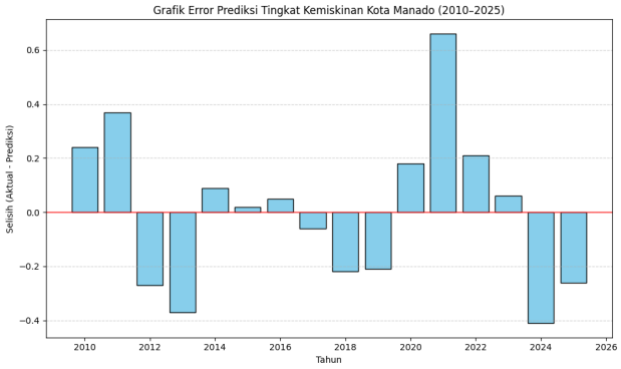
Tabel 7 menyajikan kecocokan titik-demi-titik antara nilai aktual dan nilai prediksi per tahun. Kesalahan terbesar tampak pada 2021, konsisten dengan guncangan era pandemi yang tidak sepenuhnya tercermin dalam pola historis. Ketiadaan pola over- atau under-prediction yang konsisten mengisyaratkan bahwa model relatif tidak bias antarperiode.

6) Visualisasi Hasil



Gambar 1. Tren Tingkat Kemiskinan Kota Manado (2010–2035)

Gambar 1, ditampilkan perbandingan antara tingkat kemiskinan aktual Kota Manado (garis biru) dan prediksi model regresi linier berganda (garis oranye) untuk periode 2010 hingga 2035. Tumpang tindih kedua kurva pada sebagian besar tahun mencerminkan nilai MAPE yang rendah; deviasi yang lebih lebar menandai tahun-tahun dengan guncangan idiosinkratik.



Gambar 2. Errors Tahunan (Aktual − Prediksi)

7) Implikasi Substansif dan Arah Kebijakan

Pertama, hasil menegaskan pendidikan sebagai tuas kebijakan berdaya ungkit tinggi: peningkatan rata-rata lama sekolah (X_2) berkaitan dengan penurunan tingkat kemiskinan, selaras dengan teori modal manusia. Secara praktis, ini mengarah pada program penyelesaian jenjang yang terarah di SMP/SMA, pendidikan kesempatan kedua bagi pembelajar dewasa, serta jembatan keterampilan (literasi–numerasi–digital) yang selaras dengan kebutuhan pasar kerja lokal. Peningkatan kualitas pembangunan manusia (X_3) akses kesehatan, capaian belajar, dan standar hidup perlu diposisikan sebagai masukan yang saling melengkapi, bukan pengganti pendidikan, agar capaian pada satu dimensi tidak diimbangi oleh defisit pada dimensi lain..

Kedua, agar intervensi lebih presisi, diagnosis kebijakan perlu ditautkan pada indikator pasar kerja dan makro. Memasukkan pengangguran terbuka, setengah menganggur, pergeseran komposisi sektor kerja, PDRB riil per kapita, dan inflasi membantu membedakan pendorong siklikal dari struktural kemiskinan. Kontrol tren waktu sederhana (t) atau metode deret waktu tingkat awal (ARDL/ARIMAX) memungkinkan pemantauan rutin guncangan (mis. lonjakan harga) dan efek tertunda program (mis. angkatan penerima beasiswa). Secara operasional, kota dapat memulai "dashboard risiko kemiskinan" triwulanan yang melacak indikator-indikator ini, menandai penyimpangan dari tren, dan mensimulasikan skenario dampak (mis. +0,3 tahun sekolah vs peningkatan IPM 1 poin).

Ketiga, untuk menstabilkan estimasi pada sampel kecil dan menjaga pembelajaran berkelanjutan, tim kebijakan sebaiknya mengadopsi tata kelola data dan rutinitas pemutakhiran model: (i) standarkan pipeline replikatif untuk pengumpulan data dari BPS dan perangkat daerah; (ii) pra-daftarkan himpunan fitur dan aturan validasi (LOOCV atau k-fold terblok ketika horizon meluas); (iii) pertimbangkan regularisasi (Ridge/Lasso) bila fokus pada akurasi prediktif; dan (iv) dokumentasikan keputusan dalam "kartu model" (asumsi, diagnostik, keterbatasan). Ini membentuk siklus pragmatis ukur, analisis, tindak, dan estimasi ulang sehingga program pendidikan dan

pembangunan manusia dapat diprioritaskan secara iteratif pada area dengan dampak pengurangan kemiskinan terbesar.

8) *Keterbatasan dan Pekerjaan Lanjutan*

Ukuran sampel tahunan yang kecil membatasi kompleksitas model dan kekuatan uji statistik. Penelitian selanjutnya disarankan memperpanjang horizon waktu, menambah variabel relevan, serta mengevaluasi model deret waktu (ARDL/ARIMAX) atau menerapkan regularisasi bila fokus pada prediksi murni.

Pemeriksaan deskriptif sebaiknya dilengkapi analisis korelasi Pearson antarvariabel untuk menilai arah awal keterkaitan. Jika $|r|$yang tinggi teramati antara X_2dan X_3, hal ini menegaskan perlunya evaluasi VIF sebelum menarik inferensi efek parsial.

Dengan sampel tahunan kecil, interpretasi koefisien perlu mempertimbangkan stabilitas terhadap subset waktu. Uji sensitivitas seperti mengecualikan tahun pandemi atau menambahkan kontrol tren waktu t dapat memperbaiki konsistensi arah koefisien.

LOOCV cocok untuk memaksimalkan penggunaan observasi terbatas; namun, seiring meluasnya horizon waktu, skema k-fold terblok dapat dipertimbangkan untuk menjaga ketergantungan temporal dan mengurangi kebocoran informasi.

IV. KESIMPULAN

Penelitian ini bertujuan untuk mengembangkan model regresi linier berganda untuk memprediksi tingkat kemiskinan Kota Manado (Y) dan mengevaluasi pengaruh populasi (X_1), rata-rata tahun pendidikan (X_2), dan Indeks Pembangunan Manusia (X_3). Estimasi OLS menunjukkan kesesuaian model yang memadai ($R^2 \approx 0.42$) dengan tingkat akurasi prediksi yang tinggi untuk seri waktu tahunan (MAE $\approx$ 0.23; MAPE $\approx$ 4.17%). Pemeriksaan diagnostik residu juga menunjukkan bahwa model berada dalam batas yang wajar, meskipun terdeteksi autokorelasi ringan (DW = 1.22), sehingga disarankan untuk menggunakan kesalahan standar robust (HC3) atau model berbasis tren waktu alternatif.

Untuk tujuan ini, studi ini mengestimasi Model B, yang mencakup variabel waktu sebagai pengendali tren. Hasilnya menunjukkan peningkatan stabilitas model melalui peningkatan nilai Durbin-Watson menjadi 1.69 serta peningkatan R^2 yang disesuaikan. Temuan ini menunjukkan bahwa dinamika temporal memainkan peran penting dalam memodelkan kemiskinan tahunan, dan model dengan pengendali tren waktu lebih sesuai untuk pemantauan dan prediksi kebijakan jangka panjang.

Secara substansial, variabel pendidikan (X_2) terbukti memiliki peran terkuat dalam mengurangi kemiskinan, diikuti oleh kualitas pembangunan manusia (X_3), sementara efek populasi (X_1) relatif kecil pada skala tahunan. Namun, multikolinearitas tinggi antara X_2 dan X_3 tetap menjadi batasan penting yang memerlukan kehati-hatian dalam menafsirkan koefisien.

Dari perspektif kebijakan, peningkatan rata-rata tahun pendidikan sebesar 0,5 tahun berpotensi mengurangi kemiskinan sebesar 0,10-0,12 poin persentase, sementara peningkatan Indeks Pembangunan Manusia (IPM) sebesar 1 poin dapat mengurangi kemiskinan sebesar 0,03-0,05 poin persentase. Oleh karena itu, pemerintah Kota Manado perlu memprioritaskan perluasan akses ke pendidikan menengah, mempercepat penyelesaian pendidikan wajib 12 tahun, dan meningkatkan kualitas layanan kesehatan serta produktivitas penduduk. Temuan ini dapat menjadi dasar untuk mengembangkan dashboard analitis tahunan guna memantau efektivitas kebijakan secara berkelanjutan.

Kontribusi studi ini terletak pada penyediaan model prediktif yang dapat direplikasi untuk sampel tahunan kecil, integrasi uji diagnostik model, dan perbandingan dua spesifikasi (Model A dan Model B). Penelitian lebih lanjut disarankan untuk menambahkan variabel makroekonomi, memperluas cakupan data, atau menggunakan pendekatan time-series canggih (ARDL/ARIMAX) untuk meningkatkan akurasi jangka panjang.

UCAPAN TERIMA KASIH

Penulis menyampaikan terima kasih kepada BPS Kota Manado atas dukungan akses dataset resmi dan panduan definisi. Penulis juga berterima kasih kepada rekan-rekan yang membantu dalam penyuntingan bahasa dan proofreading draf awal, serta para penelaah anonim atas masukan konstruktif yang meningkatkan kejelasan naskah ini. Segala kekeliruan yang tersisa sepenuhnya menjadi tanggung jawab penulis.

DAFTAR PUSTAKA

[1] B. K. Manado, *Indikator Kemiskinan Kota Manado Tahun 2024*, vol. 1, no. 1. 2024. [Online]. Available: http://www.biblioteca.pucminas.br/teses/Educacao_PereiraAS_1.pdf%0Ahttp://www.anpocs.org.br/portal/publicacoes/rbcs_00_11/rbcs11_01.htm%0Ahttp://repositorio.ipea.gov.br/bitstream/11058/7845/1/td_2306.pdf%0Ahttps://direitoufma2010.files.wordpress.com/2010/03/emi

[2] B. Kota Manado, "Percentage of Poor People (P0) - Kota Manado," Badan Pusat Statistik. Accessed: Nov. 06, 2025. [Online]. Available: https://manadokota.bps.go.id/en/statistics-table/2/MTIyIzI%3D/percentage-of-poor-people-p0-.html

[3] B. Kota Manado, "Indeks Kedalaman Kemiskinan (P1) di Kota Manado, 2025," BPS. Accessed: Nov. 06, 2025. [Online]. Available: https://manadokota.bps.go.id/id/statistics-table/2/MTE5IzI=/indeks-kedalaman-kemiskinan-p1-di-kota-manado.html

[4] B. Kota Manado, "Rata-rata Lama Sekolah Kota Manado, 2024," BPS. Accessed: Nov. 06, 2025. [Online]. Available: https://manadokota.bps.go.id/id/statistics-table/2/MTcyIzI=/rata-rata-lama-sekolah-kota-manado.html

[5] B. Kota Manado, "Persentase Penduduk Miskin (P0) Menurut Kabupaten/Kota (Persen), 2024," BPS. Accessed: Nov. 06, 2025. [Online]. Available: https://www.bps.go.id/id/statistics-table/2/NjIxIzI%3D/persentase-penduduk-miskin-menurut-kabupaten-kota.html

[6] BPS Sulut, "Profil Kemiskinan Provinsi Sulawesi Utara Maret 2024 - Badan Pusat Statistik Provinsi Sulawesi Utara," *Badan Pus. Stat.*, no. 42, 2025, [Online]. Available: https://sulut.bps.go.id/id/pressrelease/2024/07/01/1474/profil-kemiskinan-provinsi-sulawesi-utara-maret-2024.html

[7] N. laras Agustina, "Perkembangan Indeks Pembangunan Manusia," *Badan Pus. Stat.*, pp. 1–9, 2019.

[8] B. Kota Manado, "[New Method] Human Development Index by

Province, 2024," BPS. Accessed: Nov. 06, 2025. [Online]. Available: https://www.bps.go.id/en/statistics-table/2/NDk0IzI%3D/-new-method--human-development-index-by-province.html

[9] United Nations Development Programme, *REPORT 2023 / 2024 Reimagining cooperation in a polarized world*. 2024.

[10] Factsheet, "The World Bank's Updated Global Poverty Lines: Indonesia," The World Bank Group. Accessed: Nov. 06, 2025. [Online]. Available: https://www.worldbank.org/en/news/factsheet/2025/06/13/updated-global-poverty-lines-indonesia

[11] E. Brief and P. Rate, "Poverty & Equity Brief Poverty at Different Lines," pp. 3–4, 2024.

[12] Bappenas, "SDGs Indonesia Repository (Goal 1)," Bappenas. Accessed: Nov. 06, 2025. [Online]. Available: https://sdgs.bappenas.go.id/repository/

[13] World Bank Group, "Indonesia Country Overview," WBG. Accessed: Nov. 06, 2025. [Online]. Available: https://www.worldbank.org/en/country/indonesia/overview

[14] B. Kota Manado, "Indeks Kedalaman Kemiskinan (P1) Menurut Kabupaten/Kota," Badan Pusat Statistik. Accessed: Nov. 06, 2025. [Online]. Available: https://www.bps.go.id/id/statistics-table/2/NjIyIzI%3D/indeks-kedalaman-kemiskinan--p1--menurut-kabupaten-kota.html

[15] B. Kota Manado, "[Metode Baru] Rata-rata Lama Sekolah (Tahun), 2025," Badan Pusat Statistik. Accessed: Nov. 06, 2025. [Online]. Available: https://www.bps.go.id/id/statistics-table/2/NDE1IzI%3D/-metode-baru-rata-rata-lama-sekolah.html

[16] B. Zhao, P. K. Cheah, and P. Moses, *Measurement and determinants of multidimensional urban poverty: Evidence from Shandong Province, China*, vol. 19, no. 5 May. 2024. doi: 10.1371/journal.pone.0300263.

[17] I. Gómez-Méndez and C. Amornbunchornvej, "Income, education, and other poverty-related variables: A journey through Bayesian hierarchical models," *Heliyon*, vol. 10, no. 6, p. e27968, 2024, doi: 10.1016/j.heliyon.2024.e27968.

[18] M. Jones, "Introducing Reproducible Research Standards at the World Bank," *Harvard Data Sci. Rev.*, vol. 6, no. 4, pp. 1–9, 2024, doi: 10.1162/99608f92.21328ce3.

[19] C. Fan, M. Chen, X. Wang, J. Wang, and B. Huang, "A Review on Data Preprocessing Techniques Toward Efficient and Reliable Knowledge Discovery From Building Operational Data," *Front. Energy Res.*, vol. 9, no. March, pp. 1–17, 2021, doi: 10.3389/fenrg.2021.652801.

[20] M. Stusek *et al.*, "Accuracy Assessment and Cross-Validation of LPWAN Propagation Models in Urban Scenarios," *IEEE Access*, vol. 8, pp. 154625–154636, 2020, doi: 10.1109/ACCESS.2020.3016042.