

Pemodelan Faktor-Faktor Kemiskinan di Indonesia Menggunakan Pendekatan Regresi LASSO

Widya Febriani¹, Dony Permana²

^{1,2}Universitas Negeri Padang

¹ wfebriani978@gmail.com

ABSTRAK

Pada September 2024, persentase penduduk miskin di Indonesia mencapai 8,57 persen. Angka ini menunjukkan penurunan sebesar 0,46 persen poin dibandingkan Maret 2024. Namun capaian angka kemiskinan tersebut masih jauh dibandingkan target nasional. Penelitian ini bertujuan untuk menganalisis faktor-faktor yang mempengaruhi persentase penduduk miskin (P_0) agar dapat menjadi acuan bagi pemerintah untuk mengoptimalkan kebijakan pada faktor dominan yang memiliki pengaruh kuat terhadap persentase penduduk miskin (P_0). Metode penelitian ini menggunakan regresi LASSO dikarenakan terjadinya masalah multikolinearitas yang tinggi pada data yang digunakan. Hasil analisis menunjukkan bahwa variabel tingkat kedalaman kemiskinan (P_1) dan tingkat keparahan kemiskinan (P_2) memiliki pengaruh yang dominan terhadap persentase penduduk miskin (P_0). Nilai koefisien determinasi sebesar 95% pada data *training* dan 96% pada data *testing* menunjukkan kemampuan model sangat baik dalam menjelaskan persentase penduduk miskin (P_0).

Kata Kunci: Kemiskinan, Regresi Lasso, Kedalaman Kemiskinan, Keparahannya Kemiskinan, Indonesia

ABSTRACT

In September 2024, the percentage of poor people in Indonesia reached 8.57 percent. This figure shows a decrease of 0.46 percentage points compared to March 2024. However, this poverty rate is still far from the national target. This study aims to analyze the factors that influence the percentage of poor people (P_0) in order to provide a reference for the government to optimize policies on the dominant factors that have a strong impact on the percentage of poor people (P_0). The research method uses LASSO regression due to the high multicollinearity problem in the data used. The analysis results show that the poverty depth variable (P_1) and poverty severity variable (P_2) have a dominant influence on the percentage of poor people (P_0). The coefficient of determination value of 95% on the training data and 96% on the testing data indicates that the model has a very good ability to explain the percentage of poor people (P_0).

Keywords: Poverty, Lasso Regression, Poverty Depth, Poverty Severity, Indonesia

PENDAHULUAN

Fenomena kemiskinan di suatu negara masih menjadi permasalahan hingga saat ini. Indonesia sebagai negara dengan jumlah penduduk terbanyak ke-4 di dunia, pastinya tidak luput dari permasalahan kemiskinan. Pada September 2024, persentase penduduk miskin di Indonesia mencapai 8,57 persen. Angka ini menunjukkan penurunan sebesar 0,46 persen poin dibandingkan Maret 2024 (Badan Pusat Statistik, 2025). Meskipun terjadi penurunan, angka persentase kemiskinan di Indonesia tahun 2024 masih jauh dari target nasional. Dalam Nota Keuangan RAPBN Tahun 2024, target persentase angka kemiskinan nasional ditetapkan sebesar 6,5-7,5 persen. Adanya perbedaan yang besar antara target dengan capaian persentase kemiskinan tersebut, menyebabkan disparitas capaian angka kemiskinan antar provinsi maupun kabupaten/kota masih cukup tinggi. Permasalahan tersebut

menyebabkan analisis terkait kemiskinan penting untuk dilakukan. Pendekatan statistik yang mampu menangkap peran berbagai indikator kemiskinan sangat diperlukan agar mencapai temuan terbaru untuk membantu pemerintah dalam menghadapi masalah kemiskinan.

Dalam beberapa penelitian, pemodelan kemiskinan umumnya menggunakan pendekatan *Ordinary Least Squared* (OLS). Meskipun metode OLS banyak digunakan dalam pemodelan suatu kasus, namun metode ini memiliki keterbatasan dalam menangani kondisi dimana variable penjelas relatif banyak dan saling berkorelasi (multikolinearitas). OLS memerlukan asumsi klasik yang tidak boleh dilanggar, salah satunya adalah asumsi multikolinearitas data. Jika OLS tetap digunakan pada data yang memiliki multikolinearitas tinggi, maka hasil estimasi parameter dan inferensi di bawah prosedur OLS akan tidak signifikan dan tidak andal (Shariff & Duzan, 2018).

Perkembangan metode statistik menawarkan solusi dari permasalahan klasik tersebut. Salah satu solusi dalam mengatasi multikolinearitas dalam data yaitu melalui pendekatan regresi *Least Absolute Shrinkage and Selection Operator* (LASSO). Regresi LASSO, yang diperkenalkan oleh Robert Tibshirani pada tahun 1996, merupakan model yang dapat melakukan penyusutan pada koefisien regresi hingga mendekati atau sama dengan nol, sehingga dapat dijadikan sebagai pemilihan variabel yang relevan dan dapat mengatasi masalah multikolinearitas dalam data (Mahalani et al., 2022).

Penerapan regresi LASSO dalam kasus kemiskinan menjadi relevan untuk dilakukan. Banyaknya indikator yang digunakan dalam merepresentasikan kemiskinan, membuat peluang terjadinya masalah multikolinearitas dalam data. Beberapa penelitian terdahulu telah berhasil menerapkan metode ini terutama pada kasus kemiskinan. Penelitian yang dilakukan oleh (Darnius et al., 2024) telah membuktikan bahwa regresi LASSO mampu mengatasi masalah multikolinearitas dalam data terutama data spasial pada kasus kemiskinan yang memiliki multikolinearitas yang tinggi. Metode LASSO mengecilkan koefisien parameter yang kontribusinya lebih sedikit dan memiliki korelasi yang kuat. Selain itu, penelitian yang dilakukan oleh (Usman et al., 2021) juga membuktikan bahwa metode dengan regresi LASSO lebih unggul dibandingkan dengan metode OLS. Metode OLS gagal karena adanya multikolinearitas, sementara metode LASSO yang diregulasi menghasilkan hasil yang menunjukkan variabel prediktor sebagian besar signifikan.

Oleh karena itu, penelitian ini bertujuan untuk melihat faktor-faktor yang mempengaruhi kemiskinan menurut kabupaten/kota di Indonesia tahun 2024. Metode yang digunakan adalah analisis regresi LASSO yang lebih efektif terhadap masalah multikolinearitas dalam data.

METODE PENELITIAN

Penelitian ini merupakan penelitian eksperimen. Data penelitian yang dikumpulkan merupakan data sekunder yang diperoleh dari situs resmi Badan Pusat Statistik Indonesia tahun 2024. Cakupan wilayah penelitian terdiri dari kabupaten/kota di seluruh Indonesia dengan jumlah observasi sebanyak 514 observasi. Rincian variabel yang digunakan disajikan dalam Tabel 1 di bawah ini :

Tabel 1. Rincian Variabel Penelitian

| Nama Variabel | Penamaan | Satuan |
|------------------------------|----------|------------|
| Variabel Y | | |
| Persentasi Penduduk Miskin | P_0 | Persen (%) |
| Variabel X | | |
| Tingkat Kedalaman Kemiskinan | P_1 | Persen (%) |

| | | |
|---|----------------|--------------------|
| Tingkat Keparahan Kemiskinan | P ₂ | Persen (%) |
| Tingkat Partisipasi Angkatan Kerja | TPAK | Persen (%) |
| Produk Domestik Regional Bruto | PDRB | Milyar Rupiah (Rp) |
| Pengeluaran per Kapita | Pengeluaran | Ribu Rupiah (Rp) |
| Tingkat Pengangguran Terbuka | TPT | Persen (%) |
| Rata-Rata Lama Sekolah | RLS | Tahun |
| Persentase Rumah Tangga dengan Akses Sanitasi Layak | Sanitasi | Persen (%) |
| Usia Harapan Hidup | UHH | Tahun |
| Indeks Pembangunan Manusia | IPM | Indeks (Poin) |

Preprocessing Data

Preprocessing data mencakup berbagai operasi seperti *fitur scaling*, imputasi nilai yang hilang, pengkodean kategorikal, pemilihan fitur, dan pengurangan dimensi. Masing-masing teknik ini mengatasi tantangan spesifik dalam mempersiapkan dataset untuk analisis (Yasodha, 2025). *Preprocessing* data yang akan dilakukan dalam penelitian ini meliputi pengecekan terhadap nilai NA dan *cleaning* data seperti memastikan nilai data bertipe numerik sehingga dapat dilakukan analisis.

Splitting Data

Splitting data bertujuan untuk melatih model dapat mempelajari pola dari data historis (data *training*), lalu dievaluasi menggunakan data baru yang belum pernah digunakan dalam data pelatihan (data *testing*). Pembagian dataset penelitian ini dilakukan dengan membagi data menjadi rasio 80:20 dimana data *training* sebanyak 80% dan sisanya 20% untuk data *testing*. Pembagian ini dikarenakan rasio 80:20 dipandang ideal cukup untuk membentuk model yang seimbang (Munsarif et al., 2025).

Standarisasi Data

Setelah dilakukan *splitting* pada data, maka tahap selanjutnya yaitu standarisasi data. Standarisasi dilakukan setelah *split* data dengan tujuan untuk memastikan bahwa transformasi statistik dipelajari dari data *training*, sehingga tidak melibatkan informasi dari data *training*. Hal ini dilakukan untuk mencegah *leakage*, yaitu situasi di mana parameter transformasi dihasilkan dari keseluruhan data, sehingga berdampak pada objektivitas penilaian model (Munsarif et al., 2025).

Teknik standarisasi yang diaplikasikan dalam penelitian ini menggunakan *StandardScaler*, yaitu teknik standarisasi yang mentransformasi setiap karakteristik numerik menjadi distribusi yang memiliki nilai *mean* nol dan simpangan baku 1. Transformasi dinyatakan melalui persamaan berikut :

$$z = \frac{x - \mu}{\sigma}$$

Dimana, μ adalah rata-rata; x adalah nilai observasi dan σ adalah simpangan baku.

Multikolinearitas

Adanya multikolinearitas dalam data akan membuat model regresi yang dihasilkan menjadi tidak signifikan. Salah satu alat statistik yang dapat mendeteksi adanya multikolinearitas yaitu dengan menggunakan *Varians Inflation Factor* (VIF) (Nasution & Pane, 2024). Perhitungan nilai VIF ditunjukkan pada persamaan berikut:

$$VIF_j = C_{jj} = (1 - R_j^2)^{-1}$$

Jika nilai VIF melebihi 5 atau 10, maka adanya indikasi multikolinearitas dan sebaliknya.

Regresi LASSO

Regresi LASSO meningkatkan model regresi linier dengan memasukkan data tambahan ke dalam model melalui proses yang disebut regularisasi L1, yang mencegah *overfitting*. Oleh karena itu, dengan memasang model menggunakan setiap prediktor yang mungkin dan menggunakan teknik regularisasi yang menurunkan perkiraan koefisien menjadi nol, dapat menggunakan LASSO untuk ekstraksi fitur yang efisien. Secara khusus, tujuan pengurangan mencakup, seperti pada pengaturan regresi OLS, jumlah kuadrat residu (RSS) ditambah jumlah nilai absolut dari koefisien-koefisiennya (Guha, 2024). Dalam regresi LASSO, tujuan minimasi diubah menjadi :

$$\sum_{i=1}^n \left(y_i - (\beta_0 + \sum_{j=1}^p \beta_j x_{ij}) \right)^2 + \alpha \sum_{j=1}^p |\beta_j|$$

atau

$$RSS + \alpha \sum_{j=1}^p |\beta_j|$$

Dimana, α nilainya dapat bervariasi, tergantung pada dataset yang digunakan untuk mengestimasi nilai tersebut.

Langkah analisis penelitian dengan regresi LASSO (L1) diuraikan di bawah ini :

1. Menghitung nilai α optimal menggunakan modul *GridSearchCV* dan diimplementasikan ke dalam modul *Lasso Scikit*
2. Setelah di peroleh nilai α optimal, selanjutnya melatih model regresi LASSO menggunakan α optimal dengan data *training* yang telah di standarisasi.
3. Mengestimasi koefisien regresi. Adanya *penalty* L1 membuat beberapa koefisien akan mengalami penyusutan hingga bernilai 0.
4. Efek regulasi LASSO akan ditampilkan menggunakan grafik untuk melihat bagaimana koefisien menyusut dan dieliminasi seiring meningkatnya nilai α .
5. Lakukan seleksi fitur untuk variabel dengan koefisien regresi yang bernilai 0 karena dianggap tidak memiliki kontribusi signifikan terhadap model.
6. Selanjutnya, variabel diurutkan berdasarkan besarnya pengaruh terhadap variabel dependen.
7. Evaluasi dan validasi model

Evaluasi Hasil

Evaluasi hasil dapat menggunakan perhitungan koefisien determinasi (R^2). Koefisien determinasi adalah metrik evaluasi yang mengukur seberapa baik sebuah model cocok dengan data yang diamati dan mengkuantifikasi jumlah variasi dalam data yang dapat dijelaskan oleh model (Nur et al., 2024). Rumus perhitungan koefisien determinasi adalah sebagai berikut :

$$R^2 = 1 - \frac{SS_E}{SS_R}$$

Dimana, SS_E adalah jumlah kuadrat *error* dan SS_R adalah jumlah kuadrat total.

Selain koefisien determinasi, evaluasi hasil dapat juga dihitung menggunakan *Mean Absolute Error* (MAE). MAE merupakan matriks evaluasi dengan menghitung rata-rata perbedaan absolut antara nilai sesungguhnya dan nilai yang diprediksi. Perhitungan nilai MAE yaitu :

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

HASIL DAN PEMBAHASAN

Eksplorasi Awal

Eksplorasi data dilakukan untuk mengetahui informasi awal dalam data penelitian. Statistik deskriptif masing-masing variabel penelitian disajikan dalam Tabel 2 di bawah ini :

Tabel 2. Statistik Deskriptif Variabel

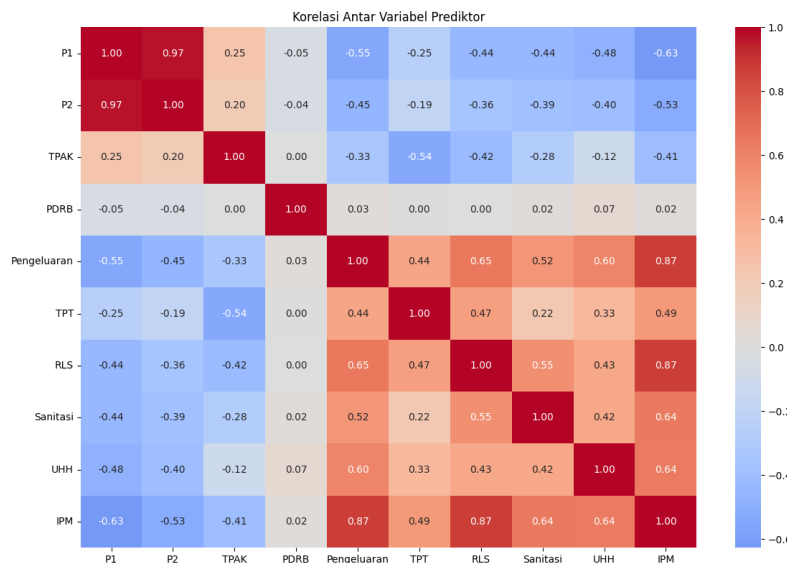
| Variabel | Mean | Standar Deviasi | Minimum | Maksimum |
|----------------|-----------|-----------------|----------|------------|
| P ₀ | 11,19 | 7,14 | 2,23 | 41,42 |
| P ₁ | 1,76 | 1,60 | 0,13 | 12,81 |
| P ₂ | 0,44 | 0,55 | 0,01 | 5,35 |
| TPAK | 71,46 | 6,15 | 55,18 | 94,81 |
| PDRB | 48.285,19 | 496.729,90 | 173,48 | 11.216.500 |
| Pengeluaran | 11.435,44 | 2.820,51 | 4.597,00 | 25.573,00 |
| TPT | 4,13 | 1,94 | 0,07 | 11,44 |
| RLS | 8,76 | 1,61 | 1,92 | 13,10 |
| Sanitasi | 80,84 | 17,45 | 2,08 | 99,68 |
| UHH | 70,41 | 3,39 | 55,74 | 78,26 |
| IPM | 73,21 | 6,17 | 36,00 | 89,10 |

Berdasarkan Tabel 2, dapat dilihat bahwa pada variabel dependent (Y) yaitu persentase penduduk miskin, memiliki nilai minimum dan maksimum yang cukup lebar. Terdapat provinsi dengan persentase penduduk miskin yang berada pada angka 41,41%, sedangkan disisi lain ada provinsi yang hanya memiliki persentase penduduk miskin sebanyak 2,23%. Hal ini mengindikasikan bahwa terdapat perbedaan yang cukup signifikan antara provinsi di Indonesia. Selain itu, pada variabel independen, hampir keseluruhan variabel memiliki nilai standar deviasi < mean, sehingga dapat dikatakan bahwa sebagian besar data relatif homogen dan tidak jauh dari rata-rata. Namun, terdapat beberapa variabel yang menunjukkan perbedaan yang jauh antara nilai minimum dan maksimum yang dihasilkan. Misalnya, pada variabel rumah tangga dengan akses sanitasi layak memiliki nilai minimum sebesar 2,23% dan maksimum sebesar 99,68%. Artinya, masih terdapat provinsi dengan akses sanitasi layak yang sangat rendah, yaitu hanya 2,23%. Sementara itu, terdapat provinsi dengan akses sanitasi layak hampir merata hingga mencapai 99,68%. Hal ini menunjukkan ketidakmerataan beberapa aspek penting seperti rumah tangga yang memiliki dan menggunakan fasilitas sanitasi (jamban/toilet) yang memenuhi standar kesehatan.

Identifikasi Multikolinearitas

Karakteristik variabel yang berbeda dan tidak seimbang dalam data penelitian dapat berpotensi menimbulkan masalah multikolinearitas. Hasil identifikasi multikolinearitas

dilakukan dengan melihat *matrix correlation* dan nilai VIF masing-masing variabel. *Matrix correlation* variabel independen ditampilkan pada Gambar 1 di bawah ini :



Gambar 1. *Matrix Correlation*

Hasil korelasi pada Gambar 1 menunjukkan bahwa terdapat beberapa variabel yang memiliki nilai korelasi yang sangat kuat. Variabel dengan nilai korelasi yang sangat kuat diantaranya seperti P₁ dengan P₂ (korelasi 0,97), IPM dan Pengeluaran (korelasi 0,87), RLS dan IPM (korelasi 0,87). Variabel lainnya juga menunjukkan nilai korelasi yang kuat seperti pengeluaran dan UHH (korelasi 0,60), pengeluaran dan RLS (korelasi 0,65), sanitasi dan IPM (korelasi 0,64). Beberapa variabel memiliki korelasi yang kuat sehingga dapat dikatakan terjadinya multikolinieritas dalam data.

Perhitungan lainnya dilakukan dengan menghitung nilai VIF untuk memastikan indikasi terjadinya multikolinieritas dalam data. Nilai VIF disajikan pada Tabel 3 di bawah ini :

Tabel 3. Statistik Deskriptif Variabel

| Variabel | Nilai VIF | Keterangan |
|----------------|-----------|---------------------------|
| P ₁ | 47,15 | Terjadi Multikolinieritas |
| P ₂ | 31,11 | Terjadi Multikolinieritas |
| TPAK | 197,68 | Terjadi Multikolinieritas |
| PDRB | 1,02 | Tidak Multikolinieritas |
| Pengeluaran | 62,86 | Terjadi Multikolinieritas |
| TPT | 9,54 | Tidak Multikolinieritas |
| RLS | 152,12 | Terjadi Multikolinieritas |
| Sanitasi | 39,41 | Terjadi Multikolinieritas |
| UHH | 764,49 | Terjadi Multikolinieritas |
| IPM | 1535,95 | Terjadi Multikolinieritas |

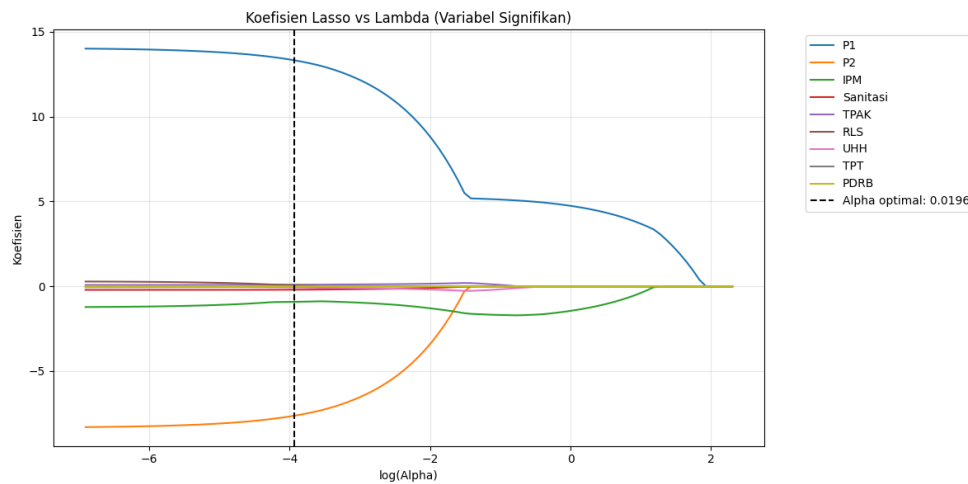
Hasil nilai VIF pada Tabel 3 membuktikan bahwa terdapat masalah multikolinieritas yang tinggi pada variabel penelitian. Hanya 2 variabel yang dinilai tidak memiliki masalah multikolinieritas. Oleh karena itu, analisis menggunakan regresi LASSO dapat dilakukan.

Hasil Analisis Regresi LASSO

Tahap awal analisis yaitu menentukan nilai α optimal. Hasil perhitungan nilai α optimal ditampilkan pada Tabel 4 :

| Kategori | Nilai |
|-------------------|-------|
| α Optimal | 0,02 |
| Log Lamda Optimal | -3,93 |

Nilai α optimal yang akan digunakan adalah 0,02. Tahap selanjutnya yaitu membuat model Lasso dengan α optimal. Dari model yang telah terbentuk, akan diidentifikasi nilai koefisien Lasso. Ketika variabel memiliki nilai koefisien = 0, maka akan dilakukan tahap seleksi variabel, sehingga variabel tersebut akan dihapus. Efek regulasi LASSO untuk melihat penyusutan koefisien ditampilkan pada Gambar 2 di bawah ini :

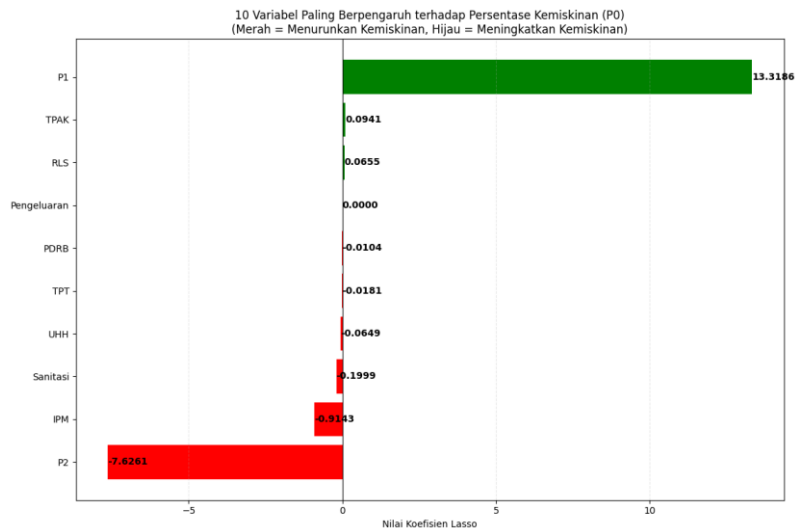


Gambar 2. Efek Regulasi Lasso

Gambar 2 menunjukkan bahwa semakin besar nilai $\log(\alpha)$ maka beberapa variabel mulai menyusut dan mendekati koefisien 0, yang menunjukkan efektivitas penalti (L1) dalam melakukan seleksi variabel. Untuk perhitungan yang lebih jelas, seleksi variabel ditampilkan pada Tabel 5 :

| Variabel | Koefisien |
|----------------|-----------|
| P ₁ | 13,32 |
| P ₂ | -7,63 |
| TPAK | 0,09 |
| PDRB | -0,01 |
| Pengeluaran | 0,00 |
| TPT | -0,02 |
| RLS | 0,07 |
| Sanitasi | 0,09 |
| UHH | 0,06 |
| IPM | -0,91 |

Berdasarkan hasil pada Tabel 5, terlihat bahwa variabel pengeluaran memiliki nilai koefisien lasso = 0, sehingga variabel pengeluaran akan dihapus dan tidak dimasukkan dalam model. Setelah tahap seleksi variabel, Langkah selanjutnya adalah mengurutkan dan mengidentifikasi variabel yang paling berpengaruh terhadap kemiskinan. Urutan variabel paling berpengaruh terhadap persentase penduduk miskin ditampilkan pada Gambar 3 berikut :

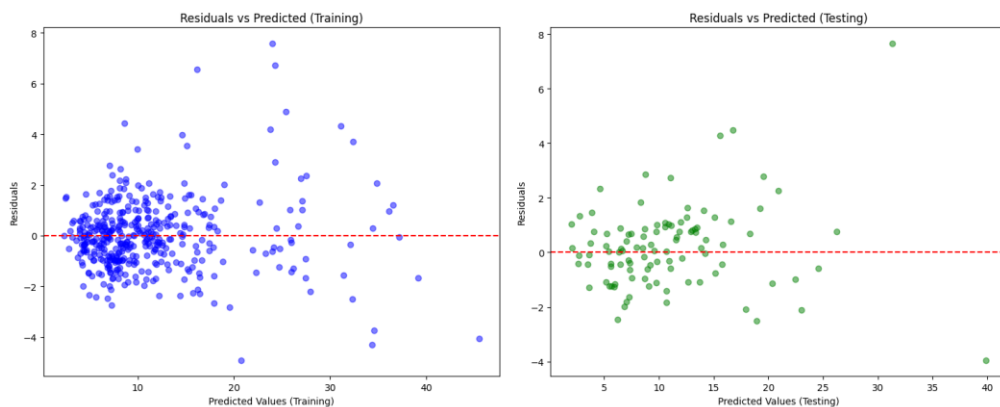


Gambar 3. Top 10 Variabel Berpengaruh Terhadap Kemiskinan

Pada Gambar 3, variabel yang paling berpengaruh adalah P₁ dengan arah pengaruh positif (meningkatkan kemiskinan), disusul oleh P₂ dengan arah pengaruh negatif (menurunkan kemiskinan). Selain itu, variabel yang memiliki arah pengaruh positif secara berurutan adalah TPAK, RLS, sedangkan variabel yang memiliki arah pengaruh negatif di antaranya IPM, Sanitasi, UHH, TPT dan PDRB.

Validasi dan Evaluasi Hasil

Validasi hasil dapat dilakukan dengan melihat pola sebaran data residual dengan data prediksi. Apabila sebaran data tidak menunjukkan / membentuk pola tertentu, maka model regresi LASSO memiliki kemampuan generalisasi yang baik. Plot sebaran data ditunjukkan pada Gambar 4 di bawah :



Gambar 4. Plot Residual vs Prediksi Data *Training* dan *Testing*

Berdasarkan Gambar 4, data residual dan prediksi pada data training maupun testing tidak menunjukkan pola sebaran tertentu. Artinya, residual menyebar secara acak di sekitar nilai nol. Hasil tersebut menunjukkan bahwa model regresi LASSO bekerja secara stabil dan estimasi yang dihasilkan dapat dipercaya.

Selanjutnya, evaluasi model dilakukan dengan melihat nilai R^2 dan MAE seperti yang disajikan pada Tabel 6 di bawah ini :

| Data | R^2 | MAE |
|-----------------|-------|------|
| <i>Training</i> | 0,96 | 1,01 |
| <i>Testing</i> | 0,95 | 1,09 |

Koefisien determinasi pada Tabel 6 menunjukkan hasil yang tinggi. Pada data *training*, koefisien determinasi sebesar 96% dan pada data *testing* sebesar 95%. Artinya, model mampu menjelaskan 95-95% informasi yang terkandung dalam data persentase penduduk miskin berdasarkan 9 variabel independen yang digunakan. Selain itu, nilai MAE sebesar 1,01 pada data *testing* dan 1,09 pada data *training* menunjukkan bahwa rata-rata kesalahan prediksi model ± 1 satuan dari nilai aktual. Besarnya informasi yang dapat dijelaskan oleh model dan kecilnya kesalahan prediksi membuktikan bahwa model yang dihasilkan dari regresi LASSO sangat baik dan stabil.

Interpretasi Hasil

Analisis menggunakan metode LASSO, menunjukkan hasil yang sangat baik terutama pada data multikolinieritas tinggi. Temuan ini sejalan dengan (Nasution & Pane, 2024) yang menyebutkan bahwa penggunaan regresi LASSO merupakan pendekatan yang lebih efektif dibandingkan dengan *Principal Component Regression* (PCR) dalam mengatasi masalah multikolinieritas pada berbagai tingkat korelasi, ukuran sampel, dan variasi variabel independen yang berbeda, karena mampu menghasilkan nilai *Mean Square Error* (MSE) yang lebih rendah serta nilai koefisien determinasi yang lebih tinggi. Lasso mengurangi masalah multikolinieritas dengan cara memilih satu variabel dari sekelompok prediktor yang sangat berkorelasi sementara variabel lainnya diperkecil menjadi nol. Dengan mengurangi kompleksitas model dan menghilangkan fitur yang tidak relevan, Lasso meminimalkan *overfitting*, memastikan kinerja yang lebih baik pada data yang belum pernah dilihat sebelumnya (Chellai, 1996).

Selain itu, dari hasil analisis diperoleh variabel yang sangat berpengaruh terhadap persentase penduduk miskin adalah tingkat kedalaman kemiskinan (P_1) yang meningkatkan persentase penduduk miskin dan tingkat keparahan kemiskinan (P_2) yang menurunkan persentase penduduk miskin. Menurut penelitian yang dilakukan oleh (Mansur, 2024) menjelaskan bahwa kenaikan indeks kedalaman kemiskinan mengindikasikan pendapatan perkapita dari penduduk miskin tidak meningkat signifikan dan tidak diimbangi dengan laju kenaikan garis kemiskinan. Hal ini mengindikasikan bahwa gap pendapatan perkapita antar sesama penduduk miskin semakin melebar, sehingga persentase penduduk miskin semakin tinggi. Namun, tingkat keparahan kemiskinan (P_2) memberi pengaruh negatif pada persentase masyarakat yang berada dalam situasi miskin (P_0). Hal ini memperlihatkan bahwa semakin tingginya keparahan kemiskinan tidak selalu diikuti kenaikan total masyarakat miskin. Dengan kata lain, meskipun kemiskinan menjadi lebih tinggi, namun

tidak menyebar ke wilayah yang lebih luas. fenomena tersebut dapat terjadi dikarenakan orang yang sudah sangat jauh di bawah batas kemiskinan mengalami kemerosotan hidup yang lebih besar, sedangkan jumlah orang yang dekat batas kemiskinan tetap cukup stabil (Celindoni, 2015).

Faktor lainnya yang memiliki pengaruh negatif cukup kuat terhadap persentase penduduk miskin adalah sanitasi dan IPM. Berkurangnya tingkat kemiskinan karena IPM yang meningkat mengindikasikan bahwa IPM dapat meningkatkan produktivitas kerja manusia, yang akan meningkatkan pendapatan untuk memenuhi kebutuhan hidup layak, sehingga menurunkan angka kemiskinan (Alvia et al., 2024). Dalam penelitian yang melibatkan beberapa negara di Asia Tenggara, juga ditemukan bahwa negara-negara dengan akses air dan sanitasi yang lebih baik cenderung mengalami tingkat kemiskinan yang lebih rendah dan lebih stabil (Putri et al., 2025). Faktor lainnya seperti UHH, TPAK, RLS, TPT dan PDRB juga turut memiliki pengaruh terhadap persentase penduduk miskin, namun pengaruh yang diberikan cukup lemah.

PEMBAHASAN

Hasil analisis menunjukkan variabel yang signifikan berpengaruh terhadap persentase penduduk miskin. Dari 10 variabel independen yang digunakan, hanya variabel pengeluaran per kapita yang tidak berpengaruh terhadap persentase penduduk miskin. Sebaliknya, variabel yang dinilai paling berpengaruh terhadap persentase kemiskinan adalah tingkat kedalaman kemiskinan (P_1) yang meningkatkan persentase penduduk miskin dan tingkat keparahan kemiskinan (P_2) yang menurunkan persentase penduduk miskin. Selain itu, variabel yang memiliki arah pengaruh positif secara berurutan adalah TPAK, RLS, sedangkan variabel yang memiliki arah pengaruh negatif di antaranya IPM, Sanitasi, UHH, TPT dan PDRB. Nilai koefisien determinasi sebesar 95%-96% dan nilai MAE sebesar 1,01-1,09 membuktikan bahwa model yang terbentuk mampu menjelaskan dengan baik persentase penduduk miskin dengan kesalahan prediksi yang rendah.

Berdasarkan hasil penelitian, diharapkan pemerintah untuk lebih berfokus pada penurunan jumlah penduduk miskin dan pengurangan kedalaman kemiskinan. Variabel yang mengukur keparahan kemiskinan memiliki dampak paling kuat terhadap persentase penduduk miskin. Variabel lain seperti IPM dan sanitasi layak juga perlu dipertimbangkan dalam menyusun kebijakan penurunan persentase kemiskinan. Oleh karena itu, pemerintah perlu memperkuat kapasitas ekonomi masyarakat terutama pada keluarga-keluarga miskin agar melewati batas kemiskinan serta meningkatkan kualitas SDM. Penelitian ini memiliki keterbatasan dimana tidak mempertimbangkan efek lokasi antawilayah. Sehingga, penelitian lanjutan diharapkan dapat mempertimbangkan efek lokasi antarwilayah seperti menggunakan regresi spasial.

UCAPAN TERIMAKASIH

Penulis mengucapkan terima kasih kepada Departemen Statistika, Universitas Negeri Padang atas dukungan fasilitas, bantuan akademik, dan lingkungan ilmiah yang mendukung kelancaran pelaksanaan penelitian ini.

REFERENSI

- Alvia, D. O., Efendi, I. W., Malihah, K. P., Kurniawan, M., Islam, U., Raden, N., Lampung, I., & Lampung, K. B. (2024). Pengaruh Indeks Pembangunan Manusia dan Pengangguran Terhadap Tingkat Kemiskinan di Indonesia Tahun 2014-2023. *Jurnal Media Akademik*, 2(6).
- Badan Pusat Statistik. (2025). *Profil Kemiskinan di Indonesia September 2024* (Badan Pusat Statistik (ed.); Issue 07). Badan Pusat Statistik.
- Celindoni, M. (2015). Decomposing Vulnerability To Poverty. *Review of Income and Wealth*, 1, 59–74. <https://doi.org/10.1111/roiw.12074>
- Chellai, F. (1996). *Lasso Regression Model*. 0–2.
- Darnius, O., Greace, Y., Turnip, C., Tarigan, E. D., & Marpaung, T. J. (2024). *Mathematical Modelling of Engineering Problems Poverty Modeling in North Sumatera Province Considering County Location Using Geographical Weighted Regression and LASSO*. 11(8), 2100–2108.
- Guha, S. (2024). Title : Feature Selection Using Lasso Regression Enhances Deep Learning Model Performance For Diagnosis Of Lung Cancer from Transcriptomic Data 1 Centre Central University), New Delhi-110025 , India Feature Selection Using Lasso Regression Enhances Deep . *BioRxiv*, 1–20.
- Mahalani, A. J., Azizah, N., Rifai, K., Statistika, P., Matematika, F., & Alam, P. (2022). Least Absolute Shrinkage And Selection Operator (LASSO) untuk Mengatasi Multikolinieritas pada Model Regresi Linear Berganda. *Bandung Conference Series: Statistics*, 2(2), 119–125.
- Mansur, Y. (2024). Analisis Perkembangan Penduduk Miskin , Karakteristik Kemiskinan dan Kedalaman Kemiskinan di Indonesia. *Jurnal Ekonomi Dan Manajemen Teknologi (EMT) KITA*, 8(1), 18–31.
- Munsarif, M. H., Fahmi, A., & Winarno, A. (2025). Evaluasi Komparatif Model Regresi Prediksi Saham BBKA dengan Analitik dan Visualisasi Interaktif Menggunakan Streamlit. *Jurnal Sistem Komputer Dan Informatika*, 7, 536–548. <https://doi.org/10.30865/json.v7i2.9119>
- Nasution, E. F., & Pane, R. (2024). Perbandingan Regresi LASSO dan Principal Component Regression dalam Mengatasi Masalah Multikolinieritas. *Journal of Mathematics Education and Science*, 10(1).
- Nur, A. R., Jaya, A. K., & Siswanto, S. (2024). Comparative Analysis of Ridge , LASSO , and Elastic Net Regularization Approaches in Handling Multicollinearity for Infant Mortality Data in South Sulawesi. *Jurnal Matematika, Statistika Dan Komputasi*, 20(2), 311–319. <https://doi.org/10.20956/j.v20i2.31632>
- Putri, B. D., Samsuddin, M. A., Studi, P., Ekonomi, I., Ekonomi, F., & Belitung, U. B. (2025). Pengaruh Akses Air Minum Layak dan Akses Sanitasi Layak terhadap Tingkat Kemiskinan : Studi Panel Kabupaten di Jawa Barat Tahun 2020- dipengaruhi oleh berbagai faktor yang saling terkait . Salah satu faktor yang diduga turut minum dan sanitasi dapat memberikan dampak ekonomi yang signifikan . *Oyebode* (2024). 3, 140–156.
- Shariff, N., & Duzan, H. (2018). A Comparison of OLS and Ridge Regression Methods in the Presence of Multicollinearity Problem in the Data. *International Journal of Engineering & Technology*, 7, 36–38.
- Usman, M., Doguwa, S. I. ., & Alhaji, B. . (2021). Comparing The Prediction Accuracy of Ridge, LASSO and Elastic Net Regression Models with Linear Regression Using Breast Cancer Data. *Bayero Journal of Pure and Applied Sciences*, 14(2), 134–149.
- Yasodha, P. (2025). Data Preprocessing Methods for Machine Learning: An Empirical Comparison. *International Journal for Multidisciplinary Research*, 7(3), 1–7.