

BAB I PENDAHULUAN

1.1. Latar Belakang

Kemiskinan merupakan salah satu permasalahan utama yang menjadi fokus perhatian pemerintah Indonesia karena berdampak pada kondisi sosial ekonomi hingga stabilitas politik suatu wilayah (Martha et al., 2021). Kurangnya pendapatan dan aset di Indonesia merupakan salah satu penyebab dari kemiskinan (*lack of income and assets*) untuk memenuhi kebutuhan dasar, seperti makanan, pakaian, tempat tinggal, kesehatan serta pendidikan (Aulina & Mirtawati, 2021). Selain itu, kemiskinan juga dapat disebabkan oleh rendahnya tingkat pendidikan, keterbatasan akses terhadap lapangan pekerjaan, kesenjangan sosial, distribusi pendapatan yang tidak merata, kondisi kesehatan yang buruk, serta dampak bencana alam atau krisis ekonomi. Menurut Badan Pusat Statistik (BPS), seseorang atau keluarga dikategorikan miskin apabila tidak mampu memenuhi kebutuhan gizi dan kebutuhan dasar lainnya. Penduduk yang memiliki pengeluaran lebih rendah dari garis kemiskinan yang telah ditetapkan disebut sebagai penduduk miskin (Soefi, 2024). Badan Pusat Statistik (2023) menginformasikan garis kemiskinan Indonesia tahun 2023 mencapai Rp550.459,- per kapita per bulan dengan jumlah persentase penduduk miskin mencapai 9,36%. Data persentase penduduk miskin yang diperoleh dari Department of Statistics Malaysia (2023) mencatat sebesar 6,20%, sementara National Statistical Office Thailand (2024) melaporkan 3,41%, dan National Statistics Office of Vietnam (2023) mencatat 3,40%. Angka ini menunjukkan bahwa tingkat kemiskinan di Indonesia masih relatif tinggi jika dibandingkan negara Asia Tenggara lainnya. Oleh karena itu, penting untuk mengidentifikasi faktor-faktor yang memengaruhi persentase penduduk miskin di Indonesia. Salah satu metode statistik yang dapat digunakan adalah analisis regresi.

Analisis regresi merupakan metode statistik yang digunakan untuk mengkaji hubungan antara satu atau lebih variabel prediktor terhadap variabel respon (Septa et al., 2025). Pembentukan model regresi dilakukan dengan menduga koefisien regresi yang berperan sebagai estimasi parameter dalam model tersebut. Estimasi parameter regresi umumnya menggunakan metode *Ordinary Least Squares* (OLS) (Gunawan et al., 2024). Metode OLS memiliki beberapa asumsi dasar yang harus dipenuhi agar estimasi parameter regresi bersifat tidak bias dan efisien. Salah satu asumsi penting tersebut adalah tidak adanya multikolinearitas, yaitu kondisi variabel prediktor tidak memiliki korelasi linear yang tinggi satu sama lain. Namun, apabila jumlah variabel prediktor cukup banyak atau terjadi korelasi tinggi antar variabel,

menghasilkan model yang kurang optimal (Mait et al., 2021). Oleh karena itu, *SSO Regression* dapat digunakan sebagai alternatif (Kamila et

al., 2021). *Lasso Shrinkage and Selection Operator* (LASSO) Regression adalah metode regularisasi yang digunakan dalam model regresi untuk mengatasi masalah multikolinearitas dan *overfitting* dengan menambahkan penalti



pada fungsi objektif (Najib et al., 2024). Model ini dapat menyusutkan koefisien regresi pada variabel prediktor yang memiliki korelasi dan galat yang tinggi, dengan tujuan koefisien regresi tersebut mendekati nol atau bahkan sama dengan nol (Sartika et al., 2020). Adapun kelebihan dari *LASSO Regression* adalah kemampuannya dalam melakukan seleksi variabel prediktor sehingga hanya variabel yang berpengaruh disertakan dalam model. Hal ini dapat mempermudah proses interpretasi terhadap model regresi yang telah dihasilkan (Pardede et al., 2022). Akan tetapi, *LASSO Regression* memiliki keterbatasan karena tidak mampu mempertimbangkan struktur spasial atau hubungan kedekatan antar objek sehingga hal ini dapat mengakibatkan hilangnya informasi mengenai keterkaitan antar variabel (R. Tibshirani et al., 2005). Oleh karena itu, model yang dihasilkan berpotensi kurang akurat dalam menggambarkan pola sebenarnya yang terjadi pada data, seperti pada data persentase penduduk miskin di Indonesia yang tersebar antar wilayah. Sebagai solusi atas keterbatasan tersebut, digunakan model *Generalized LASSO Regression* yang merupakan pengembangan dari *LASSO Regression*.

Model *Generalized LASSO Regression* dapat mengakomodasi struktur kedekatan objek dengan menambahkan sebuah matriks penalti, sehingga model dapat menyeleksi variabel sekaligus pengelompokan objek berdasarkan kedekatannya melalui pemilihan parameter *tuning* yang tepat. Pembentukan matriks penalti tersebut dapat dilakukan dengan berbagai pendekatan, salah satunya melalui algoritma *K-Nearest Neighbors* (KNN) yang mempertimbangkan jarak atau kedekatan antar wilayah (Juhanda, 2024). KNN merupakan metode yang berbasis kedekatan atau jarak yang digunakan untuk mengidentifikasi objek terdekat dari suatu titik data berdasarkan ukuran tertentu, seperti jarak *euclidean* dalam konteks spasial (Halder et al., 2024). Pada penerapan model *Generalized LASSO Regression*, KNN digunakan untuk menentukan struktur hubungan antar objek atau wilayah dengan mengidentifikasi tetangga terdekat dari masing-masing amatan wilayah.

Beberapa peneliti telah melakukan penelitian mengenai model *LASSO Regression* dengan data spasial seperti penelitian yang dilakukan oleh Kusuma & Wulansari (2019) mengenai analisis kemiskinan dan kerentanan di Provinsi Jawa Tengah tahun 2017 menghasilkan nilai *Mean Squared Error LASSO Regression* lebih kecil dari *Ridge Regression*. Penelitian yang dilakukan oleh Anwar (2022) dengan menggunakan *Geographically Weighted LASSO* (GWL) diperoleh bahwa model GWL lebih baik model *Geographically Weighted Regression* (GWR). Penelitian yang dilakukan oleh Mutaqin et al. (2024), melakukan klusterisasi menggunakan *Generalized LASSO Regression* dengan pendekatan *Queens* dan KNN, hasil dari

ukkan nilai parameter *tuning* KNN lebih optimal dibandingkan



engenai faktor-faktor yang memengaruhi kemiskinan telah ada umumnya, penelitian mengenai kemiskinan fokus pada al dan ekonomi di suatu wilayah. Menurut penelitian yang gun & Meimela (2020), ditemukan bahwa indeks pembangunan ngkat partisipasi angkatan kerja (TPAK) berpengaruh signifikan

terhadap persentase penduduk miskin di beberapa wilayah Sumatera Utara. Nizar & Arif (2023), pada penelitiannya menghasilkan pengeluaran per kapita berpengaruh negatif terhadap tingkat kemiskinan di Nusa Tenggara Barat. Selain itu, penelitian oleh Anam & Feriyanto (2024) menunjukkan hasil angka partisipasi sekolah (APS) dan IPM berpengaruh negatif terhadap kemiskinan di Kabupaten Purworejo. Berdasarkan penelitian-penelitian tersebut, model *Generalized LASSO Regression* dengan pendekatan KNN digunakan untuk mengeksplor pengaruh variabel APS, pengeluaran riil per kapita, IPM, dan TPAK terhadap persentase penduduk miskin.

Penggunaan metode *Generalized LASSO Regression* dengan pendekatan *K-Nearest Neighbors* (KNN) dalam penelitian ini bertujuan untuk mengeksplorasi dan mengidentifikasi variabel-variabel sosial ekonomi yang paling berpengaruh terhadap tingkat kemiskinan. Metode ini tidak hanya mampu melakukan seleksi variabel, tetapi juga efektif dalam menangani multikolinearitas antar variabel, seperti IPM, APS, pengeluaran per kapita, dan TPAK. Pendekatan KNN diterapkan untuk mempertimbangkan aspek spasial, yaitu dengan mengasumsikan bahwa wilayah-wilayah yang saling bertetangga memiliki karakteristik kemiskinan yang serupa. Dengan demikian, model ini dapat memberikan hasil yang akurat dan relevan secara geografis. Selain itu, penalti dalam *Generalized LASSO Regression* membantu menyederhanakan model, mengurangi risiko *overfitting*, dan meningkatkan kemampuan prediksi terhadap tingkat kemiskinan di wilayah lain.

1.2. Batasan Masalah

Batasan masalah pada penelitian ini adalah sebagai berikut:

1. Data yang digunakan dalam penelitian ini merupakan data persentase penduduk miskin pada seluruh provinsi di Indonesia pada tahun 2023.
2. Penelitian ini menggunakan model *Generalized LASSO Regression* dengan matriks ketetanggaan *K-Nearest Neighbors*.

1.3. Tujuan dan Manfaat Penelitian

Tujuan yang ingin dicapai dalam penelitian ini adalah sebagai berikut.

1. Mendapatkan parameter *tuning* optimal dalam model *Generalized LASSO Regression* dengan pendekatan KNN.
2. Mendapatkan variabel yang memengaruhi persentase penduduk miskin pada setiap bagian amatan di Indonesia melalui model *Generalized LASSO Regression*.

Berdasarkan tujuan penelitian di atas, maka setelah penelitian ini dilakukan diharapkan dapat memberikan manfaat sebagai berikut:



• er pengetahuan bagi akademisi, peneliti dan mahasiswa yang estimasi parameter model *Generalized LASSO Regression* unakan data spasial.

• er pengetahuan dan informasi mengenai persentase penduduk setiap bagian amatan di Indonesia sehingga bermanfaat bagi

pemerintah dan pemangku kebijakan untuk meningkatkan kesejahteraan masyarakat di Indonesia.

1.4. Landasan Teori

1.4.1 Regresi Linier Berganda

Penelitian yang mengkaji pengaruh antara variabel prediktor (X) dan variabel respon (Y) umumnya menggunakan pendekatan analisis statistik, salah satunya adalah dengan menggunakan regresi linier. Regresi linier digunakan untuk mengetahui seberapa besar pengaruh perubahan variabel prediktor terhadap variabel respon. Selain itu, regresi linier juga dapat memberikan informasi mengenai pengaruh hubungan variabel respon dan variabel prediktor, dengan pengaruh searah atau berlawanan arah. Bentuk umum model regresi linier sederhana adalah sebagai berikut:

$$Y = \beta_0 + \beta X + \varepsilon$$

dengan β_0 adalah nilai intersep model regresi, β adalah estimasi koefisien regresi pengamatan variabel X , dan ε adalah *error*.

Pengembangan dari regresi linier sederhana adalah regresi linier berganda. Pendekatan ini digunakan untuk menganalisis hubungan antara dua atau lebih variabel prediktor dan satu variabel respon. Bentuk umum regresi linier sederhana dan nilai *error* dapat diperoleh menggunakan Persamaan (1) dan (2) berikut (Darma, 2021):

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p + \varepsilon \quad (1)$$

$$\begin{aligned} \varepsilon &= (Y - \hat{Y}) \\ &= (Y - X\hat{\beta}) \end{aligned} \quad (2)$$

dengan β_1 adalah estimasi koefisien regresi pengamatan variabel X_1 , dan p adalah jumlah variabel prediktor.

Menurut Setiyorini et al. (2017), nilai estimasi koefisien dalam regresi linier dapat dihitung menggunakan *Ordinary Least Squares* (OLS) dengan meminimalkan nilai jumlah kuadrat *error* (ε^2) antara variabel respon aktual (Y) dan variabel respon prediksi ($\hat{Y}$). Proses ini bertujuan untuk memperoleh koefisien model yang meminimalkan perbedaan antara nilai amatan dan nilai prediksi, sehingga dapat menghasilkan nilai estimasi koefisien yang mendekati nilai amatan. Nilai ε^2 dapat dihitung menggunakan Persamaan (3) berikut:



$$\begin{aligned} \varepsilon^2 &= \varepsilon^T \varepsilon \\ \varepsilon^T \varepsilon &= (Y - X\hat{\beta})^T (Y - X\hat{\beta}) \\ &= Y^T Y - Y^T X\hat{\beta} - YX^T \hat{\beta}^T + X^T \hat{\beta}^T X\hat{\beta} \\ &= Y^T Y - 2YX^T \hat{\beta}^T + X^T \hat{\beta}^T X\hat{\beta} \\ &= -2X^T Y + 2X^T X\hat{\beta} \\ &= -X^T Y + X^T X\hat{\beta} \end{aligned} \quad (3)$$

Koefisien regresi β diperoleh melalui OLS pada Persamaan (3) dan menghasilkan Persamaan (4) sebagai berikut:

$$\hat{\beta} = (X^T X)^{-1} X^T Y \quad (4)$$

dengan Y adalah vektor variabel respon Y , $\hat{\beta}$ adalah vektor estimasi koefisien regresi, X adalah matriks variabel prediktor berukuran $n \times p$, dan n adalah jumlah amatan variabel respon Y .

1.4.2 Asumsi Normalitas

Asumsi normalitas adalah asumsi statistik yang dirancang untuk menilai sekumpulan data mengikuti distribusi normal atau tidak. Salah satu uji normalitas yang sering digunakan adalah uji *Shapiro-Wilk*. Uji normalitas *Shapiro-Wilk* digunakan pada data dengan jumlah pengamatan n yang tidak terlalu besar. Uji normalitas *Shapiro-Wilk* dapat dilihat pada Persamaan (5) berikut:

$$W = \frac{(\sum a_i (Y_{n-i+1} - Y_i))^2}{\sum (Y_i - \bar{Y})^2} \quad (5)$$

dengan Y_i adalah variabel respon baris ke i , $\bar{Y}$ adalah rata-rata variabel respon. Sebuah data dapat dikatakan berdistribusi normal jika memenuhi hipotesis dan kriteria pengujian normalitas.

Hipotesis:

H_0 : Data berdistribusi normal

H_1 : Data tidak berdistribusi normal

Pengambilan keputusan dalam uji *Shapiro-Wilk* didasarkan pada perbandingan antara nilai *p-value* dan taraf signifikansi α . Hipotesis nol (H_0) diterima jika $p - value \geq \alpha$, sedangkan H_0 ditolak saat $p - value < \alpha$. Pada umumnya, nilai α ditetapkan sebesar 5% atau sebesar 0,05 (Khatun, 2021).

1.4.3 Asumsi Multikolinearitas

Asumsi multikolinearitas pada model regresi dapat terjadi Ketika dua atau lebih variabel prediktor memiliki nilai korelasi yang tinggi satu sama lain. Korelasi yang tinggi antar variabel prediktor berdampak pada nilai β yang tidak stabil. Pendeteksian multikolinearitas dapat diperoleh dengan menghitung nilai *Variance Inflation Factor* (VIF). Nilai VIF masing-masing variabel dapat dihitung menggunakan Persamaan (6) berikut:

$$VIF_m = \frac{1}{1 - R_m^2} \quad (6)$$

variabel ke- m , dan R_m^2 adalah koefisien determinasi variabel > 10 , maka terjadi multikolinearitas pada model regresi.



Menurut Shrestha (2020), uji korelasi *Pearson* dapat digunakan untuk mengidentifikasi korelasi antar pasangan variabel. Uji korelasi *Pearson* memiliki nilai koefisien dengan rentang $[-1,1]$, apabila nilai koefisien korelasi semakin mendekati salah satu batas tersebut maka hubungan linier antar variabel semakin kuat. Uji korelasi *Pearson* dapat dihitung menggunakan Persamaan (7) berikut:

$$r_m = \frac{n \sum X_m Y - \sum X_m \sum Y}{\sqrt{(n \sum X_m^2 - (\sum X_m)^2)(n \sum Y^2 - (\sum Y)^2)}} \quad (7)$$

dengan r_m adalah koefisien korelasi *Pearson* variabel ke- m . Jika hasil uji korelasi *Pearson* menghasilkan nilai $r \leq 0,40$ maka korelasi antar variabel yang terjadi lemah, sedangkan jika nilai $0,40 < r \leq 0,60$ maka terjadi korelasi sedang antar variabel, dan jika $r \leq 0,61$ maka terjadi korelasi kuat hingga sempurna antar variabel (Hidayanti et al., 2023). Nilai korelasi positif menunjukkan hubungan searah yang berarti ketika variabel respon meningkat maka variabel prediktor juga meningkat. Nilai korelasi negatif menunjukkan hubungan berlawanan yang berarti ketika variabel respon meningkat maka variabel prediktor menurun.

1.4.4 Least Absolute Shrinkage and Selection Operator (LASSO) Regression

LASSO Regression merupakan metode regresi untuk mengatasi multikolinearitas dengan cara mereduksi koefisien parameter β menjadi tepat pada nol atau mendekati nol sehingga menghasilkan parameter β yang lebih sedikit dan model yang dihasilkan lebih mudah diinterpretasikan dibandingkan dengan model awal. Pada LASSO Regression, sebuah penalti ditambahkan ke dalam Persamaan (2). Adapun estimasi koefisien LASSO Regression dituliskan pada Persamaan (8) berikut (Rusdy et al., 2025) :

$$\hat{\beta}^{lasso} = \arg \min_{\beta} \left\{ \frac{1}{2} \|Y - X\beta\|_2^2 + \lambda \|\beta\|_1 \right\} \quad (8)$$

dengan $\lambda \|\beta\|_1$ merupakan penalti LASSO Regression (L_1), dan λ merupakan nilai parameter *tuning* optimal.

Model LASSO Regression mendorong koefisien regresi menjadi lebih kecil bahkan menjadi nol lalu membentuk variabel prediktor terpilih. Model ini akan mempertahankan hanya variabel prediktor yang memberikan kontribusi dan secara otomatis menyinkronkan variabel prediktor non-kontributif.

1.4.5 Generalized LASSO Regression



Taylor (2011), *Generalized LASSO Regression* menerapkan sien seperti LASSO Regression. Pengembangan *Generalized* terdapat pada penambahan matriks penalti D sehingga pada Persamaan (8). Matriks D bertujuan untuk bla struktur vektor β . Saat koefisien β berubah secara bertahap spasial) atau waktu (temporal), matriks D dibentuk sebagai

matriks perbedaan berurutan. Penalti L_1 yang diterapkan kemudian menyatukan nilai koefisien β ke dalam tiap blok sehingga menjadi konstan, sehingga pola perubahan yang diharapkan tampak dalam hasil estimasi. Estimasi koefisien dengan pemodelan *Generalized LASSO Regression* dapat ditunjukkan pada Persamaan (9) berikut:

$$\hat{\beta}^{genlasso} = \arg \min_{\beta} \left\{ \frac{1}{2} \|Y - X\beta\|_2^2 + \lambda \|D\beta\|_1 \right\} \quad (9)$$

dengan D adalah matriks penalti yang ditentukan, seperti hubungan spasial. Penelitian yang dilakukan oleh Rahardianto & Sakamoto (2021), menjelaskan bahwa *Generalized LASSO Regression* dapat digunakan untuk data yang berbentuk spasial untuk mengestimasi koefisien berdasarkan spesifik wilayah. Berikut estimasi koefisien *Generalized LASSO Regression* berdasarkan spesifik wilayah yang ditunjukkan pada Persamaan (10):

$$\min_{\beta} \left\{ \sum (Y_{lk} - \beta_{l0} - X_{lk1}\beta_{l1} - X_{lk2}\beta_{l2} - X_{lk3}\beta_{l3} - X_{lk4}\beta_{l4})^2 + \lambda \|D\beta\|_1 \right\} \quad (10)$$

dengan X_{lk1} , X_{lk2} , X_{lk3} dan X_{lk4} adalah variabel prediktor, $l = 1, 2, 3, \dots, 6$ mewakili 6 bagian, dan $k = 1, 2, \dots$ menunjukkan provinsi ke- k pada tiap bagian. Matriks D dibentuk berdasarkan kedekatan antar 6 bagian. Koefisien regresi masing-masing bagian dipertimbangkan sehingga membentuk $X \in \mathbb{R}^{n \times p}$ dan $D \in \mathbb{R}^{m \times p}$ dengan m adalah jumlah hubungan ketetanggaan yang terbentuk.

1.4.6 K-Nearest Neighbors

K-Nearest Neighbors (KNN) merupakan algoritma *machine learning* yang diterapkan pada analisis klasifikasi dan regresi pada sebuah penelitian statistika. Prinsip kerja KNN adalah mencari sejumlah K tetangga terdekat dari data yang akan diprediksi. KNN dikenal sebagai algoritma *machine learning* yang efektif untuk klasifikasi dan regresi. Algoritma ini berdasar pada asumsi bahwa objek-objek dalam radius tertentu (bertetangga) yaitu objek data yang memiliki lokasi atau posisi yang saling berdekatan berdasarkan ukuran jarak tertentu (biasanya *Euclidean distance*) akan memiliki pola atau karakteristik yang serupa. Penentuan estimasi nilai untuk objek baru dilakukan dengan mengevaluasi tingkat kemiripan antara objek tersebut dan tetangga-tetangga terdekatnya (Ernawati et al., 2025).

Algoritma KNN digunakan dalam menentukan K tetangga terdekat untuk setiap amatan wilayah dengan mempertimbangkan titik koordinat geografisnya (*longitude, latitude*). Penerapan KNN menggunakan metrik jarak untuk menentukan



wilayah amatan dengan menghitung jarak titik koordinat setiap natan. Pada umumnya, metrik jarak yang digunakan adalah *euclidean*. Pada Persamaan (11) berikut dijabarkan ri metrik jarak *euclidean*:

$$d_{u,v} = \sqrt{(Long_u - Long_v)^2 + (Lat_u - Lat_v)^2} \quad (11)$$

dengan $d_{u,v}$ adalah jarak *euclidean* antar wilayah u dan v , $Long_u, Lat_u$ adalah koordinat *longitude* dan *latitude* wilayah ke- u , dan $Long_v, Lat_v$ adalah koordinat *longitude* dan *latitude* wilayah ke- v . Pada amatan wilayah dengan gabungan beberapa wilayah, pengambilan titik *longitude* dan *latitude* amatan dilakukan dengan menentukan titik *centroid* dari kumpulan beberapa wilayah amatan yang tergabung dalam satu amatan. *Centroid* merupakan titik pusat atau rata-rata posisi dari semua gabungan wilayah amatan. Adapun titik *centroid* dapat ditentukan melalui Persamaan (12) dan (13) berikut:

$$Long_s = \frac{1}{N} \sum_{i=1}^N Long_c \quad (12)$$

$$Lat_s = \frac{1}{N} \sum_{i=1}^N Lat_c \quad (13)$$

dengan N merupakan jumlah wilayah amatan yang digabungkan, $Long_c$ dan Lat_c merupakan jumlah titik koordinat *longitude* dan *latitude* gabungan.

1.4.7 Matriks D Menggunakan *K-Nearest Neighbors*

Matriks penalti D pada *Generalized LASSO Regression* dirancang untuk menangkap struktur keterkaitan, salah satunya keterkaitan spasial antar wilayah pengamatan. Matriks penalti dasar D_0 dibangun terlebih dahulu menggunakan algoritma KNN dalam menentukan hubungan ketetanggaan spasial antar wilayah pengamatan. Algoritma KNN menentukan setiap lokasi u yang dihubungkan dengan K tetangga terdekatnya.

Matriks D_0 yang terbentuk mewakili setiap pasang tetangga (u, v) pada barisnya, dengan nilai pada kolom u bernilai 1, kolom v bernilai -1 , sedangkan kolom lainnya diisi dengan 0 (Gong et al., 2017). Sehingga, perkalian $D_0\beta$ menghasilkan selisih koefisien $\beta_u - \beta_v$ untuk setiap pasang tetangga. Pendekatan ini memetakan perbedaan koefisien antara lokasi-lokasi yang bertetangga secara geografis (Rahardiantoro et al., 2024).

Matriks D_0 diperluas ke dalam format *block-diagonal*, sehingga setiap *block* mewakili hubungan ketetanggaan antar wilayah pengamatan berdasarkan K tetangga yang dipertimbangkan. Pembuatan matriks identitas (I) bertujuan untuk menerapkan konsep *self-adjacency* dengan menempatkan secara diagonal entri bernilai 1. Konsep *self-adjacency* digunakan untuk mempresentasikan hubungan tiap wilayah amatan dengan dirinya sendiri. Matriks ini kemudian dipasangkan secara $\ll D_0$, sehingga membentuk matriks akhir D .



ross Validation

tuning (λ) optimal pada model *Generalized LASSO Regression* menggunakan metode *Cross Validation*. Menurut λ optimal yang diidentifikasi oleh model *Approximate Leave-*

One-Out Cross Validation (ALOCV) dan *Generalized Cross Validation* (GCV) memberikan hasil yang lebih efisien dibandingkan *K-Fold Cross Validation*. Ketika nilai λ cukup kecil, ALOCV tidak dapat diterapkan karena dapat menyebabkan model tidak stabil. Berbeda dengan ALOCV, model GCV dapat mengatasi situasi tersebut dan menghasilkan estimasi parameter *Generalized LASSO Regression* yang optimal. Kriteria GCV dapat dirumuskan pada Persamaan (14) sebagai berikut (Jansen, 2015):

$$GCV = \frac{1}{n \sum_{i=1}^n (Y_i - \hat{Y}_i)^2} \left(1 - \frac{df_{\lambda}}{n}\right)^{-2} \quad (14)$$

dengan $\hat{Y}_i$ merupakan nilai Y_i estimasi dari persamaan regresi, dan df_{λ} merupakan derajat kebebasan model pada λ .

1.4.9 Uji Kelayakan Model

Root Mean Squared Error (RMSE) dan koefisien determinasi (R^2) digunakan sebagai pengujian kelayakan pada model regresi, seperti *Generalized LASSO Regression*. RMSE memberikan informasi tentang kinerja suatu model yang dapat mengevaluasi performa model melalui perbandingan perbedaan aktual antara nilai estimasi dan nilai yang diukur. Semakin kecil nilai RMSE, maka semakin baik pula kinerja model. R^2 digunakan untuk mengukur proporsi variansi target yang berhasil dijelaskan oleh model, dengan rentang [0,1]. Nilai R^2 yang mendekati 1 menginterpretasikan seberapa besar model dalam menjelaskan data. Nilai RMSE dan R^2 dapat dihitung melalui Persamaan (15) dan (16) berikut (Yildirim, 2024):

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2} \quad (15)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} \quad (16)$$

1.4.10 Persentase Penduduk Miskin

Kemiskinan merupakan salah satu masalah dan tantangan besar dalam pembangunan nasional suatu negara. Meskipun berbagai upaya telah dilakukan oleh pemerintah, namun kemiskinan masih menjadi faktor utama dalam ketimpangan



masyarakat. Pada negara berkembang seperti Indonesia, ia mencerminkan ketidakmampuan individu atau rumah tangga memenuhi kebutuhan dasar, tetapi juga mencerminkan ketimpangan akses, seperti pendidikan, lapangan pekerjaan, serta partisipasi ekonomi (Fitri et al., 2024).

Penduduk miskin menjadi indikator awal untuk melihat masalah yang ada di suatu wilayah. Persentase penduduk miskin menunjukkan

proporsi individu atau sebuah keluarga dengan pengeluaran yang berada dibawah garis kemiskinan. Berdasarkan laporan BPS (2023), persentase penduduk miskin Indonesia mencapai 9,36% atau sekitar 25,90 juta jiwa. Jika dibandingkan dengan periode yang sama pada tahun sebelumnya angka persentase penduduk miskin ini mengalami penurunan sebesar 0,18% atau sekitar 0,46 juta jiwa. Akan tetapi, penurunan tersebut tidak terjadi secara merata di seluruh wilayah Indonesia. Tingkat kemiskinan di Indonesia masih menunjukkan perbedaan yang signifikan antar wilayah, ini menunjukkan bahwa faktor-faktor penyebab tingkat kemiskinan bervariasi di setiap wilayah, terutama pada pulau-pulau di Indonesia.



BAB II METODE PENELITIAN

2.1. Sumber Data dan Variabel

Data yang digunakan dalam penelitian ini merupakan data sekunder mengenai persentase penduduk miskin menurut provinsi di Indonesia pada tahun 2023 yang diperoleh dari www.bps.go.id. Jumlah amatan pada penelitian ini adalah 34 provinsi di Indonesia.

2.2. Deskripsi Variabel

Variabel yang digunakan dalam penelitian ini terdiri dari 1 variabel respon dan 4 variabel prediktor yang dapat dilihat pada Tabel 1.

Tabel 1. Variabel Penelitian

Variabel	Keterangan	Satuan
Y	Penduduk Miskin	Persentase (%)
X_1	Angka Partisipasi Sekolah 13-15 Tahun	Persentase (%)
X_2	Pengeluaran Riil per Kapita	Ribu
X_3	Indeks Pembangunan Manusia	Indeks
X_4	Tingkat Partisipasi Angkatan Kerja	Persentase (%)

2.3. Metode Analisis

Langkah-langkah analisis data yang dilakukan berdasarkan tujuan penelitian adalah sebagai berikut:

1. Melakukan eksplorasi data terhadap hubungan antar variabel
 - a. Sebaran masing-masing variabel respon dan variabel prediktor pada peta Indonesia.
 - b. Melakukan uji multikolinearitas pada variabel prediktor (X) dengan uji VIF menggunakan Persamaan (6).
 - c. Menghitung koefisien korelasi *Pearson* antar seluruh variabel prediktor terhadap variabel respon dengan menggunakan Persamaan (7).
 - d. Melakukan uji asumsi kenormalan pada variabel respon (Y) dengan uji normalitas *Shapiro-Wilk* menggunakan Persamaan (5). Jika variabel Y tidak normal, lakukan transformasi logaritmik seperti pada Persamaan (17) berikut:

$$Y_{i \text{ Log}} = \text{Log}(Y_i) \quad (17)$$

- a. Melakukan standarisasi data menggunakan z-score terhadap variabel yang dapat dilihat pada Persamaan (18) berikut:

$$Z_{ij} = \frac{x_{ij} - \bar{x}_j}{(s \times \sqrt{n-1})} \quad (18)$$

di mana $\bar{x}_j$ adalah rata-rata variabel x_j , dan s adalah standar deviasi



- f. Menyesuaikan struktur data yang dapat dilihat pada Gambar 1 dan Gambar 2 berikut:

Bagian	Provinsi	Y	X1	X2	X3	X4
A	1	Y(A1)	X1(A1)	X2(A1)	X3(A1)	X4(A1)
	2	Y(A2)	X1(A2)	X2(A2)	X3(A2)	X4(A2)
	3	Y(A3)	X1(A3)	X2(A3)	X3(A3)	X4(A3)
B	1	Y(B1)	X1(B1)	X2(B1)	X3(B1)	X4(B1)
	2	Y(B2)	X1(B2)	X2(B2)	X3(B2)	X4(B2)
	3	Y(B3)	X1(B3)	X2(B3)	X3(B3)	X4(B3)
	4	Y(B4)	X1(B4)	X2(B4)	X3(B4)	X4(B4)
	5	Y(B5)	X1(B5)	X2(B5)	X3(B5)	X4(B5)
	6	Y(B6)	X1(B6)	X2(B6)	X3(B6)	X4(B6)
C	1	Y(C1)	X1(C1)	X2(C1)	X3(C1)	X4(C1)
	2	Y(C2)	X1(C2)	X2(C2)	X3(C2)	X4(C2)
	3	Y(C3)	X1(C3)	X2(C3)	X3(C3)	X4(C3)
	4	Y(C4)	X1(C4)	X2(C4)	X3(C4)	X4(C4)
D	1	Y(D1)	X1(D1)	X2(D1)	X3(D1)	X4(D1)
	2	Y(D2)	X1(D2)	X2(D2)	X3(D2)	X4(D2)
	3	Y(D3)	X1(D3)	X2(D3)	X3(D3)	X4(D3)
	4	Y(D4)	X1(D4)	X2(D4)	X3(D4)	X4(D4)
	5	Y(D5)	X1(D5)	X2(D5)	X3(D5)	X4(D5)
	6	Y(D6)	X1(D6)	X2(D6)	X3(D6)	X4(D6)
E	1	Y(E1)	X1(E1)	X2(E1)	X3(E1)	X4(E1)
	2	Y(E2)	X1(E2)	X2(E2)	X3(E2)	X4(E2)
	3	Y(E3)	X1(E3)	X2(E3)	X3(E3)	X4(E3)
	4	Y(E4)	X1(E4)	X2(E4)	X3(E4)	X4(E4)
	5	Y(E5)	X1(E5)	X2(E5)	X3(E5)	X4(E5)
	6	Y(E6)	X1(E6)	X2(E6)	X3(E6)	X4(E6)
	7	Y(E7)	X1(E7)	X2(E7)	X3(E7)	X4(E7)
	8	Y(E8)	X1(E8)	X2(E8)	X3(E8)	X4(E8)
	9	Y(E9)	X1(E9)	X2(E9)	X3(E9)	X4(E9)
	10	Y(E10)	X1(E10)	X2(E10)	X3(E10)	X4(E10)
F	1	Y(F1)	X1(F1)	X2(F1)	X3(F1)	X4(F1)
	2	Y(F2)	X1(F2)	X2(F2)	X3(F2)	X4(F2)
	3	Y(F3)	X1(F3)	X2(F3)	X3(F3)	X4(F3)
	4	Y(F4)	X1(F4)	X2(F4)	X3(F4)	X4(F4)
	5	Y(F5)	X1(F5)	X2(F5)	X3(F5)	X4(F5)

Gambar 1. Struktur Awal

Bagian	Provinsi	Y	B0							X1			X4		
			A	B	C	D	E	F	A	..	F	..	A	..	F
A	1	Y(A1)	1	0	0	0	0	0	X1(A1)	..	0	..	X4(A1)	..	0
	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
	3	Y(A3)	1	0	0	0	0	0	X1(A3)	..	0	..	X4(A3)	..	0
B	1	Y(B1)	0	1	0	0	0	0	0	..	0	..	0	..	0
	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
	6	Y(B6)	0	1	0	0	0	0	0	..	0	..	0	..	0
C	1	Y(C1)	0	0	1	0	0	0	0	..	0	..	0	..	0
	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
	4	Y(C4)	0	0	1	0	0	0	0	..	0	..	0	..	0
D	1	Y(D1)	0	0	0	1	0	0	0	..	0	..	0	..	0
	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
	6	Y(D6)	0	0	0	1	0	0	0	..	0	..	0	..	0
E	1	Y(E1)	0	0	0	0	1	0	0	..	0	..	0	..	0
	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
F	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
G	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
H	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
I	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
J	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
K	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
L	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
M	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
N	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
O	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
P	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
Q	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
R	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
S	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
T	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
U	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
V	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
W	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
X	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
Y	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
Z	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..

Gambar 2. Struktur Baru



2. Melakukan evaluasi model.
 - a. Menyusun matriks penalti D menggunakan KNN dengan $K = 1$, $K = 2$, dan $K = 3$ sebagai dasar dalam proses *Generalized LASSO Regression* secara spasial.
 - b. Menentukan nilai λ optimal menggunakan pendekatan *Generalized Cross Validation* (GCV) dengan Persamaan (14).
 - c. Melakukan uji validasi model dengan menggunakan *Root Mean Square Error* (RMSE) dan R^2 dengan menggunakan Persamaan (15) dan Persamaan (16).
3. Menghitung nilai koefisien regresi model *Generalized LASSO Regression*.
 - a. Memodelkan data dengan *Generalized LASSO Regression* menggunakan Persamaan (10) dengan memasukkan nilai λ optimal dari proses GCV.
 - b. Menginterpretasikan hasil dan melakukan penarikan kesimpulan berdasarkan hasil analisis yang telah diperoleh.

