

BAB I

PENDAHULUAN

1.1 Latar Belakang

Indeks Pembangunan Manusia (IPM) adalah ukuran komposit yang mengevaluasi tingkat pencapaian suatu wilayah dalam tiga dimensi utama: kesehatan, pendidikan, dan standar hidup yang layak (UNDP, 2020). Sebagai indikator kunci kesejahteraan, IPM menjadi perhatian utama bagi pemerintah daerah dalam merencanakan dan mengevaluasi efektivitas kebijakan pembangunan. Provinsi Sulawesi Selatan, dengan dinamika sosial dan ekonominya yang khas, menjadi studi kasus yang menarik untuk menganalisis faktor-faktor yang memengaruhi IPM. Beberapa faktor utama yang berkontribusi terhadap IPM meliputi rata-rata lama sekolah (RLS), yang mencerminkan tingkat akses dan kualitas pendidikan. Pengeluaran per kapita, yang mencerminkan daya beli masyarakat sebagai indikator kesejahteraan ekonomi. Umur harapan hidup (UHH), yang menggambarkan kualitas layanan kesehatan dan tingkat kesejahteraan masyarakat. Tingkat pengangguran terbuka (TPT), yang menjadi indikator kondisi ketenagakerjaan dan peluang ekonomi. Pemahaman terhadap determinan IPM di Provinsi Sulawesi Selatan memiliki dampak signifikan pada berbagai aspek kehidupan masyarakat.

Analisis regresi merupakan salah satu pendekatan model statistika yang umum digunakan dalam menganalisis pengaruh satu atau beberapa variabel prediktor terhadap satu variabel respon (Johan Harlan, 2018). Metode estimasi parameter yang umum digunakan dalam analisis regresi adalah *Ordinary Least Squares* (OLS). Namun, keberadaan multikolinearitas dalam data dapat menyebabkan estimasi parameter regresi dengan metode OLS menjadi tidak efisien. Multikolinearitas disebabkan oleh hubungan linear yang kuat antar variabel prediktor dan penggunaan jumlah variabel yang relatif banyak (Sungkono & Nugrahaningsih, 2017). Menurut Aboye dkk. (2014), multikolinearitas memengaruhi standar residual dalam estimasi koefisien regresi, sehingga dapat menyebabkan kesalahan dalam interpretasi model. Hal ini dianggap sebagai kelemahan dalam analisis regresi dengan metode OLS, karena dapat menghasilkan kesimpulan yang tidak tepat (Millenia, 2021). Oleh karena itu, penting untuk mengatasi masalah multikolinearitas, salah satunya dengan menggunakan metode *Least Absolute Shrinkage and Selection Operator*.

Metode *Least Absolute Shrinkage and Selection Operator* (LASSO) merupakan teknik regresi berbasis penalti yang mampu menyusutkan nilai estimator koefisien regresi. Metode LASSO juga mampu melakukan seleksi variabel prediktor karena penalti yang diterapkan dapat mengecilkan beberapa estimator hingga bernilai nol. Estimator LASSO diperoleh dengan meminimalkan jumlah kuadrat *residual* dengan menambahkan penalti atau regularisasi yang dikenal sebagai L_1 ke dalam proses estimasi koefisien regresi. Gonzales (2022) menyatakan bahwa dalam proses LASSO, penambahan penalti atau regularisasi L_1 ke dalam model estimator OLS sebagai langkah untuk mengestimasi nilai koefisien regresi β dapat menimbulkan kelemahan, yaitu munculnya bias. Untuk mengoreksi bias tersebut,

dikembangkan versi terbobot dari metode LASSO berdasarkan skema *iteratif*. Namun, LASSO memiliki keterbatasan karena menerapkan penalti yang sama pada semua parameter, sehingga sering kali kurang optimal untuk data dengan multikolinearitas tinggi.

Pengembangan lebih lanjut dari LASSO adalah *adaptive* LASSO, yang menggunakan bobot adaptif sehingga mampu menghasilkan estimasi parameter yang akurat (Zou, 2006). Dengan adanya bobot ini, *adaptive* LASSO lebih fleksibel dalam menyusutkan variabel dan mempertahankan variabel yang signifikan dalam model (Zou & Hastie, 2005). Metode ini lebih optimal dalam menangani multikolinearitas dibandingkan LASSO karena dapat memberikan estimasi parameter yang lebih akurat serta seleksi variabel yang lebih stabil dalam model regresi, dengan proses *adaptive* LASSO yang diselesaikan menggunakan algoritma *coordinate descent* (Friedman et al., 2008).

Algoritma *coordinate descent* merupakan salah satu metode optimasi numerik yang efisien dalam mengestimasi nilai penalti dalam *adaptive* LASSO. Algoritma ini bekerja dengan cara memperbarui satu koefisien regresi pada satu waktu sementara koefisien lainnya dianggap tetap, sehingga proses iterasi menjadi lebih sederhana dibandingkan metode optimasi lainnya seperti *gradient descent*. *Coordinate descent* efektif dalam menangani model dengan jumlah variabel yang besar saat melakukan estimasi parameter pada *adaptive* LASSO. Oleh karena itu, penggunaan algoritma *coordinate descent* dalam *adaptive* LASSO memungkinkan seleksi variabel menjadi akurat, serta menghasilkan model regresi yang stabil dan mudah diinterpretasikan.

Penelitian yang dilakukan oleh Friedman (2007) menggunakan algoritma *coordinate descent* untuk menyelesaikan masalah optimisasi dalam metode LASSO. Hasil penelitian tersebut menunjukkan bahwa algoritma ini memiliki performa yang sebanding dengan prosedur lain yang sudah dikenal dalam penyelesaian masalah LASSO. Metode LASSO telah diterapkan dalam berbagai bidang, termasuk oleh Yusuf et al. (2021) yang menggunakan LASSO untuk memodelkan variabel sosial ekonomi yang memengaruhi kesejahteraan masyarakat. Hasil penelitian tersebut menunjukkan bahwa LASSO efektif dalam mengidentifikasi variabel-variabel signifikan yang berpengaruh terhadap kesejahteraan masyarakat. Pengembangan lebih lanjut dikemukakan oleh Fajrina (2017) dalam penelitian mengenai penyusutan koefisien dan akurasi model *adaptive* LASSO. Hasilnya menunjukkan bahwa *adaptive* LASSO mampu menyusutkan variabel-variabel yang tidak signifikan menjadi nol, sehingga meningkatkan akurasi dan interpretasikan model.

Berdasarkan uraian yang dibahas sebelumnya, maka penulis ingin mengkaji peran *Adaptive* LASSO dalam pemilihan konstanta penalti di lakukan secara adaptif dan menyusutkan koefisien variabel yang tidak signifikan, sehingga hanya variabel yang berpengaruh terhadap faktor-faktor IPM yang dipertahankan. Analisis data menunjukkan adanya pelanggaran asumsi multikolinearitas yang dapat memengaruhi akurasi estimasi parameter. Mengatasi hal tersebut, penelitian ini menerapkan metode *adaptive* LASSO dengan algoritma *coordinate descent* untuk menghasilkan model yang lebih stabil.

1.2 Batasan Masalah

Pada penelitian ini terdapat beberapa batasan-batasan yang akan di teliti. Batasan-batasan ini digunakan untuk mempermudah penelitian dalam melakukan suatu penelitian, yaitu:

1. Permodelan regresi linear menggunakan metode *adaptive* LASSO dengan memanfaatkan algoritma *coordinate descent* pada proses estimasi parameter koefisien regresi.
2. Objek yg diteliti yaitu indeks pertumbuhan manusia di Provinsi Sulawesi Selatan tahun 2021 yg mengandung data multikolinearitas.

1.3 Tujuan dan Manfaat Penelitian

Tujuan dari penelitian ini adalah sebagai berikut:

1. Mendapatkan dugaan model *adaptive* LASSO pada data IPM di Provinsi Sulawesi Selatan tahun 2021.
2. Memperoleh faktor-faktor yang signifikan mempengaruhi IPM di Provinsi Sulawesi Selatan tahun 2021 berdasarkan model *adaptive* LASSO.

Manfaat yang diperoleh dari penelitian ini adalah sebagai berikut:

1. Menambah ilmu dan wawasan terkait *adaptive* LASSO untuk mengatasi masalah multikolinearitas pada data IPM di Provinsi Sulawesi Selatan tahun 2021.
2. Penelitian ini diharapkan dapat memberikan masukan kepada pemerintah sebagai dasar pertimbangan bagi pembuat kebijakan ekonomi daerah untuk memanfaatkan model prediksi pertumbuhan ekonomi yang lebih akurat, sehingga dapat digunakan untuk perencanaan pembangunan di Provinsi Sulawesi Selatan.

1.4 Teori

1.4.1 Regresi linear

Regresi merupakan analisis statistik yang digunakan untuk mencari hubungan antara dua variabel atau lebih. Variabel-variabel tersebut digolongkan menjadi dua yaitu variabel Y (variabel respon) dan variabel X (variabel prediktor). Regresi disebut sebagai regresi linear apabila variabel respon dengan variabel prediktornya memiliki hubungan yang linear. Terdapat dua jenis regresi linear yaitu regresi linear sederhana dan regresi linear berganda. Bentuk umum persamaan regresi linear berganda dapat ditulis sebagai berikut (Montgomery dkk., 2013):

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_j X_{ij} + \dots + \beta_k X_{ik} + \varepsilon_i \quad (1)$$

Keterangan:

Y_i : variabel respon pada pengamatan ke- i) $i = 1, 2, \dots, n$)

β_0 : konstanta

β_j : koefisien regresi variabel prediktor ke- j ($j = 1, 2, \dots, k$)

X_{ij} : variabel prediktor ke- j untuk pengamatan ke- i

ε_i : variabel residual ke- i

Persamaan (1) dapat ditulis ke dalam notasi matriks berikut:

$$\begin{bmatrix} Y \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & \dots & X_{1k} \\ 1 & X_{21} & \dots & X_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & \dots & X_{nk} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Bentuk matriks di atas dapat ditulis secara sederhana menjadi:

$$Y = X\beta + \varepsilon \quad (2)$$

1.4.2 Metode *Ordinary Least Square*

Salah satu metode yang umum digunakan dalam mengestimasi koefisien regresi β adalah metode *Ordinary Least Square* (OLS). Metode ini meminimumkan jumlah kuadrat *error*. Dalam metode ini dibutuhkan beberapa asumsi yang perlu dipenuhi. Pemeriksaan asumsi dalam regresi mencakup beberapa aspek, yaitu linearitas, normalitas *error*, *mean error* nol, homogenitas varians, dan multikolinearitas. Linearitas dapat diperiksa dengan *scatter plot*, di mana pola garis menunjukkan hubungan linear antara variabel respon dan prediktor. Normalitas *error* dapat diuji secara visual melalui histogram, boxplot, dan pp-plot, yang menunjukkan distribusi simetris atau titik-titik yang mengikuti garis diagonal, serta dapat diuji secara statistik dengan uji *Kolmogorov-Smirnov*. *Mean error* bernilai nol jika *scatter plot* menunjukkan titik-titik yang tersebar acak di sekitar garis nol tanpa pola tertentu. Homoskedastisitas terpenuhi jika variansi *error* konstan, yang terlihat dari pola sebaran titik yang merata di sekitar garis mean = 0. Sementara itu, multikolinearitas dapat dideteksi menggunakan matriks korelasi atau nilai *Variance Inflation Factor* (VIF), di mana hubungan linear yang kuat antarvariabel prediktor menunjukkan adanya multikolinearitas.

Metode OLS mengestimasi koefisien regresi β dengan meminimumkan jumlah kuadrat *error* dengan bentuk persamaan dapat dituliskan sebagai berikut :

$$\begin{aligned} S(\beta_0, \beta_1, \dots, \beta_k) &= \sum_{i=1}^n \varepsilon_i^2 \\ &= \sum_{i=1}^n \left(Y_i - (\beta_0 + \sum_{j=1}^k \beta_j X_{ij}) \right)^2 \end{aligned} \quad (3)$$

Estimator OLS dari $\beta_0, \beta_1, \dots, \beta_k$ diperoleh dengan menurunkan fungsi S terhadap masing-masing parameter koefisien regresinya dan harus memenuhi persamaan berikut :

$$\frac{\partial S}{\partial \beta_0} \Big|_{\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k} = -2 \sum_{i=1}^n (Y_i - \hat{\beta}_0 - \sum_{j=1}^k \hat{\beta}_j X_{ij}) = 0$$

dan

$$\frac{\partial S}{\partial \beta_j} \Big|_{\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k} = -2 \sum_{i=1}^n (Y_i - \hat{\beta}_0 - \sum_{j=1}^k \hat{\beta}_j X_{ij}) X_{ij} = 0$$

Sehingga diperoleh estimator OLS dari β adalah

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

Pada regresi linear berganda, estimasi parameter koefisien regresi dilakukan dalam bentuk notasi matriks. Berdasarkan Persamaan (2), diperoleh

$$\varepsilon = Y - X\beta$$

Estimator OLS dari β dapat diestimasi dengan meminimalkan jumlahan kuadrat *error* yang ditulis sebagai fungsi berikut :

$$\begin{aligned} S(\beta) &= \sum_{i=1}^n \varepsilon_i^2 = \varepsilon^T \varepsilon \\ &= (Y - X\beta)^T (Y - X\beta) \\ &= (Y^T - \beta^T X^T) (Y - X\beta) \\ &= Y^T Y - Y^T X\beta - \beta^T X^T Y + \beta^T X^T X\beta \end{aligned} \quad (4)$$

karena $\beta^T X^T Y$ merupakan skalar, dan $(\beta^T X^T Y)^T = Y^T X\beta$ memiliki nilai skalar yang sama maka Persamaan (9) dapat dituliskan.

$$S(\beta) = Y^T Y - 2\beta^T X^T Y + \beta^T X^T X\beta \quad (5)$$

Estimator OLS diperoleh dengan menurunkan fungsi $S(\beta)$ terhadap masing-masing parameter koefisien regresinya dan harus memenuhi $\frac{\partial S(\beta)}{\partial \beta} = 0$. Maka diperoleh:

$$\begin{aligned} -2X^T Y + 2X^T X\hat{\beta} &= 0 \\ X^T X\hat{\beta} &= X^T Y \end{aligned} \quad (6)$$

Sehingga diperoleh estimator OLS dari β adalah :

$$\hat{\beta}_{OLS} = (X^T X)^{-1} X^T Y \quad (7)$$

1.4.3 Multikolinearitas

Multikolinearitas terjadi ketika dua atau lebih variabel prediktor dalam model memiliki hubungan yang kuat satu sama lain. Multikolinearitas terjadi apabila terdapat hubungan antara beberapa atau seluruh variabel prediktor. Multikolinearitas dalam suatu model regresi dapat diketahui dengan menghitung nilai *Variance Inflation Factor* (VIF). Gejala multikolinearitas terjadi jika nilai $VIF \geq 10$ menunjukkan bahwa variabel tersebut sangat berkorelasi dengan variabel lainnya. Perhitungan nilai VIF dapat dilakukan menggunakan rumus berikut (Montgomery dkk., 2013):

$$VIF_j = \frac{1}{1 - R_j^2} \quad (8)$$

Dengan R_j^2 menyatakan nilai koefisien determinasi masing-masing dari variabel prediktor X_j yang diregresikan terhadap variabel prediktor lainnya.

1.4.4 Fungsi Sign

Fungsi *sign* merupakan fungsi yang menunjukkan tanda dari suatu bilangan, atau yang dapat didefinisikan sebagai berikut:

$$\text{sign}(x) = \begin{cases} -1, & \text{jika } x < 0 \\ 0, & \text{jika } x = 0 \\ 1, & \text{jika } x > 0 \end{cases} \quad (9)$$

Fungsi *sign* sendiri merupakan hasil *derivative* dari fungsi mutlak atau *absolut* yang akan ditunjukkan sebagai berikut:

$$f(x) = |x| \quad (10)$$

Bentuk fungsi $f(x)$ dapat ditulis sebagai $f(x) = |x| = \sqrt{x^2} = (x^2)^{\frac{1}{2}}$ akan ditunjukkan bahwa hasil *dervative* dari $f(x)$ adalah *sign* (x).

$$f'(x) = \frac{\partial (x^2)^{\frac{1}{2}}}{\partial x}$$

$$\begin{aligned}
 f'(x) &= \frac{1}{2}(x^2)^{-\frac{1}{2}}(2x) \\
 &= \frac{x}{(x^2)^{\frac{1}{2}}} \\
 &= \frac{x}{|x|} = \text{sign}(x)
 \end{aligned}$$

Dengan $x \neq 0$

Sehingga untuk setiap bilangan real dinyatakan dalam bentuk berikut:

$$x = |x| \text{sign}(x)$$

1.4.5 Metode Centering dan Rescaling

Salah satu metode standarisasi pada data yang digunakan adalah dengan menggunakan metode *centering* dan *rescaling*. Menurut Kutner (2005), Standarisasi data diperlukan untuk memastikan bahwa variabel dengan skala yang berbeda tidak mendominasi analisis statistik, terutama dalam metode regresi seperti *adaptive LASSO*. Dengan melakukan *centering*, kita menghilangkan skala rata-rata dari setiap variabel sehingga interpretasi koefisien menjadi lebih mudah. *Rescaling* memastikan bahwa semua variabel memiliki skala yang sebanding, yaitu dalam satuan standar deviasi, sehingga algoritma estimasi parameter lebih stabil dan tidak terpengaruh oleh perbedaan skala antar variabel.

Berikut standarisasi yang biasa dilakukan pada variabel respon Y dan variabel prediktor X .

$$X_{ij}^* = \frac{X_{ij} - \bar{X}_j}{s_{X_j}}, \quad i = 1, 2, \dots, n; \quad j = 1, 2, \dots, k$$

Dan

$$Y_i^* = \frac{Y_i - \bar{Y}}{s_{Y_i}}, \quad i = 1, 2, \dots, n$$

Dimana

$$s_{X_j}^2 = \frac{\sum_{i=1}^n (X_{ij} - \bar{X}_j)^2}{n - 1}$$

Merupakan variasi dari variabel prediktor X_j , dan

$$s_{Y_j}^2 = \frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{n - 1}$$

Merupakan variasi dari variabel respon

Dengan melakukan standarisasi maka baik variabel respon maupun variabel prediktor akan memiliki mean nol dan variasinya 1. Dilakukannya proses standarisasi akan diperoleh variabel baru dimana persamaan regresi modelnya menjadi:

$$Y_i^* = \beta_1^* X_{i1}^* + \beta_2^* X_{i2}^* + \dots + \beta_k^* X_{ik}^* + \varepsilon_i \quad (11)$$

Untuk $i = 1, 2, \dots, n$

Centering variabel respon dari variabel prediktor dengan mengurangi nilainya dengan X dan Y akan menghilangkan *intercept* dari model pada estimasi *least square* nilai $b_0 = b = Y^* = 0$. Estimator *least square* adalah

$$\beta^* = (X^T X)^{-1} X^T Y^* \quad (12)$$

Adapun hubungan antara koefisien $\beta_1^*, \beta_2^*, \dots, \beta_{k-1}^*$ pada model regresi hasil standarisasi dan koefisien $\beta_0, \beta_1, \dots, \beta_{p-1}$ pada model regresi awal dapat ditunjukkan sebagai berikut:

$$\beta_j = \frac{S_{Yj}}{S_{Xj}} \beta_k^* \quad (j = 1, 2, \dots, k-1) \quad (13)$$

$$\beta_0 = \bar{Y} - \beta_1 \bar{X}_1 - \dots - \beta_{k-1} \bar{X}_{k-1} \quad (14)$$

1.4.6 Algoritma Coordinate Descent

Coordinate Descent Algorithm merupakan algoritma yang digunakan untuk memecahkan masalah optimasi dengan meminimumkan nilai fungsi yang dapat diperoleh dari meminimalkan arah koordinat secara berurutan. Algoritma ini telah diaplikasikan selama bertahun-tahun dan cukup populer berkat kegunaannya dalam analisis data dan *machine learning*. Friedman (2007) menyebutkan bahwa keunggulan utama coordinate descent adalah kemampuannya dalam meminimalkan fungsi secara efisien pada setiap koordinat, menangani variabel yang berulang dengan baik, serta menjalankan iterasi dengan cepat.

Algoritma *coordinate descent* merupakan metode *iterative* dimana setiap iterasi memperbarui sebagian besar komponen vektor variabel X berdasarkan nilai iterasi saat ini, dengan tujuan meminimalkan fungsi secara parsial terhadap komponen yang tersisa.

Permasalahan minimasi tak terbatas secara umum dapat dituliskan sebagai berikut:

$$\min_x f(X) = f(X_1, X_2, \dots, X_n) \quad (15)$$

Dimana $X = (X_1, X_2, \dots, X_n) \in \mathbb{R}^n$ dan fungsi $f : \mathbb{R}^n \rightarrow \mathbb{R}$ kontinu

Selain untuk bentuk diatas, terdapat pula permasalahan minimasi dengan struktur berikut:

$$\min_x F(X) = f(X) + \sum_{i=1}^n r_i(X_i) \quad (16)$$

Dimana f terdiferensiasi dan r_i merupakan *extended-valued* dan mungkin fungsinya tidak dapat terdiferensiasi. Untuk setiap nilai r_i dapat memodelkan pada X_i dengan memasukkan fungsi indikator. Fungsi r_i juga dapat memuat aturan regularisasi tertentu dalam menentukan struktur solusi.

Adapun algoritma *coordinate descent* dapat digambarkan sebagai berikut:

1. Mengukur nilai $k = 0$ dan menginisialisasi $X^0 \in \mathbb{R}^n$.
2. Secara berulang dilakukan:
 - a. Memilih indeks i dari 1 sampai n atau $i_k \in \{1, 2, \dots, n\}$;
 - b. Meng-update nilai X_{ik} ke X_{ik}^k dengan skema tertentu tergantung pada X^{k-1} dan f atau F
 - c. Biarkan X_j bernilai tetap, contohnya $X_j^k = X_j^{k-1}$ untuk setiap $j \neq i_k$
 - d. Mengeset $k = k + 1$;

Hingga tercapainya konvergensi atau kondisi terminasi terpenuhi, (Shi, 2017)

1.4.7 Metode *Least Absolute Shrinkage and Selection Operator*

Least Absolute Shrinkage and Selection Operator (LASSO) merupakan metode yang diperkenalkan oleh Tibshirani pada tahun 1996 sebagai alternatif dari metode *ridge* standar dan sebagai teknik *subset selection*. Metode ini digunakan untuk mengestimasi koefisien regresi serta menyeleksi variabel prediktor yang masuk ke dalam model regresi. LASSO mampu menyusutkan koefisien regresi dari variabel prediktor yang berkorelasi hingga nilainya mendekati nol atau bahkan sama dengan nol, sehingga dapat menghasilkan estimasi koefisien regresi dengan variasi yang lebih kecil, lebih representatif, serta mampu mengatasi masalah multikolinearitas. Adapun estimasi dari koefisien regresinya diperoleh dengan meminimumkan jumlah kuadrat *error* dengan menambahkan penalti pada prosesnya yang kemudian disebut sebagai penalti L_1 . Estimasi parameter dari model regresi linear dengan metode LASSO dapat dituliskan dalam bentuk berikut:

$$\hat{\beta}_{LASSO} = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^n \left(Y - \beta_0 - \sum_{j=1}^k X_{ij} \beta_j \right)^2 \quad (17)$$

Dengan kendala

$$\sum_{j=1}^k |\beta_j| \leq t \quad (18)$$

Keterangan:

Y_i : variabel respon ke- i

β_j : koefisien regresi dari variabel prediktor ke- j

X_{ij} : variabel prediktor ke- j pada pengamatan ke- i

n : jumlah pengamatan

k : jumlah variabel prediktor dalam model

t : parameter tuning LASSO, dengan $t \geq 0$

Diasumsikan bahwa kovariatnya telah distandarisasi sehingga memiliki rata-rata nol dan variansi satu:

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n X_{ij} &= 0 \\ \frac{1}{n} \sum_{i=1}^n X_{ij}^2 &= 1 \end{aligned}$$

Sehingga untuk semua nilai t , solusi untuk β_0 adalah $\hat{\beta}_0 = Y$. Dengan asumsi *Loss of generality*, dapat diasumsikan bahwa $Y = 0$, sehingga $\hat{\beta}_{LASSO}$

$$\hat{\beta}_{LASSO} = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^n \left(y_i - \sum_{j=1}^k X_{ij} \beta_j \right)^2$$

Dengan kendala $\sum_{j=1}^k |\beta_j| \leq t$

1.4.8 Metode Adaptive Least Absolute Shrinkage and Selection Operator

Metode *adaptive Least Absolute Shrinkage and Selection Operator* (LASSO) merupakan salah satu pengembangan dari metode LASSO yang pertama kali diperkenalkan oleh Hui Zou pada tahun 2006. Metode ini dianggap mampu memberikan kondisi optimal sehingga pemilihan variabel dalam LASSO menjadi lebih konsisten. Zou (2006) menyatakan bahwa metode LASSO menghasilkan estimasi yang bias terhadap koefisien yang besar, sehingga kurang optimal dalam proses estimasi. Untuk mengatasi permasalahan tersebut, dikembangkan versi baru dari LASSO yang disebut sebagai *adaptive LASSO*.

Menurut Ghassany (2010), dalam metode *adaptive LASSO*, diberikan bobot adaptif pada koefisien yang berbeda dalam penalti L_1 , sehingga menghasilkan fungsi penalti yang bervariasi untuk setiap koefisiennya. Dengan demikian, besarnya penyusutan untuk setiap koefisien pun berbeda. Menambahkan bobot pada persamaan LASSO, diperoleh persamaan baru yang mempresentasikan estimator *adaptive LASSO* sebagai berikut:

$$\hat{\beta}_{ALASSO} = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^n \left(Y_i - \sum_{j=1}^k X_{ij} \beta_j \right)^2 + \lambda \sum_{j=1}^k W_j |\beta_j| \quad (19)$$

di mana $W_j = (W_1, W_2, \dots, W_k)$ adalah vektor bobot yang dapat didefinisikan sebagai:

$$W_j = \frac{1}{|\beta_j|^\gamma} \quad (20)$$

dengan $\gamma > 0$, $0 \leq W < \infty$ dan β_j merupakan estimator OLS. Semakin besar nilai γ , maka semakin besar kekuatannya dalam menyusutkan nilai estimator. Apabila nilai $\gamma = 0$, maka nilai vektor bobot W_j mendekati 1, sehingga penalti *adaptive LASSO* menjadi penalti LASSO. Sedangkan, apabila $\gamma = \infty$ akan terjadi seleksi *subset* terhadap variabel prediktor.

Dalam kasus tertentu, yaitu saat $\gamma = 1$, *adaptive LASSO* dapat dianggap sebagai *nonnegative garotte*. Zou (2006) menyatakan bahwa *nonnegative garotte* dinilai konsisten. Oleh karena itu, untuk *adaptive LASSO* diambil γ yang nilainya 1, sehingga vektor bobot *adaptive* dapat dituliskan sebagai:

$$W_j = \frac{1}{|\hat{\beta}_{OLSj}|} \quad (21)$$

Dari uraian diatas maka persamaan untuk *adaptive LASSO* dapat dituliskan kembali ke dalam bentuk berikut:

$$\hat{\beta}_{ALASSO} = \underset{\beta}{\operatorname{argmin}} \left(\|Y - X\beta\|^2 + \lambda \sum_{j=1}^k W_j |\beta_j| \right)$$

Apabila dalam bentuk matriks, estimasi parameter dari model regresi linear dengan metode *adaptive LASSO* diperoleh persamaan berikut:

$$\hat{\beta}_{ALASSO} = (X^T X + \lambda W)^{-1} X^T Y$$

Dengan:

$$W = \operatorname{diag}(W_1, W, \dots, W_k)$$

Jika didefinisikan

$$\beta = W ; X = XW^{-1}$$

maka

$$\begin{aligned} S &= (Y - X\beta)^T(Y - X\beta) + \lambda W|\beta| \\ &= (Y - X\hat{\beta})^T(Y - X\hat{\beta}) + \lambda|\hat{\beta}| \\ &= Y^TY - 2\hat{\beta}^TX^TY + \hat{\beta}^TX^TX\hat{\beta} + \lambda|\hat{\beta}| \end{aligned}$$

Estimator *adaptive LASSO* diperoleh dengan menurunkan fungsi S terhadap masing-masing parameter koefisien regresinya dan harusnya memenuhi $\frac{\partial S}{\partial \beta} = 0$ maka diperoleh

$$\begin{aligned} 0 &= -2X^TY + 2X^TX\hat{\beta} + \lambda \text{sign}(\hat{\beta}) \\ X^TX\hat{\beta} &= X^TY - \frac{\lambda}{2} \text{sign}(\hat{\beta}) \\ \hat{\beta} &= (X^TX)^{-1}X^TY - \frac{\lambda}{2}(X^TX)^{-1} \text{sign}(\hat{\beta}) \\ \hat{\beta} &= \hat{\beta}_{OLS} - \frac{\lambda}{2}(X^TX)^{-1} \text{sign}(\hat{\beta}) \end{aligned}$$

dengan $\text{sign}(\hat{\beta}) = \text{sign}(\hat{\beta}_{OLS})$, dan karena matriks X dalam keadaan orthonormal, maka $X^TX = I$, sehingga persamaan dapat ditulis kembali menjadi

$$\hat{\beta} = \left(|\hat{\beta}_{OLS}| - \frac{\lambda}{2} W \right)^+ \text{sign}(\hat{\beta}_{OLS}) \quad (22)$$

1.4.9 Pemilihan Parameter *Tuning*

Pada metode seperti LASSO dan *adaptive LASSO* pemilihan parameter *tuning* atau nilai rumus akan memengaruhi besarnya penyusutan, untuk itu memilih parameter *tuning* yang tepat sangatlah penting. Salah satu metode yang umum digunakan dalam pemilihan parameter *tuning* atau nilai rumus dengan metode validasi silang (*Cross Validation*)

Adapun langkah-langkah *K-folds cross validation* menurut Tibshirani (2013) adalah sebagai berikut:

1. Membagi data set menjadi K bagian terpisah yang mempunyai ukuran sama sehingga $T = T_1, T_2, \dots, T_K$. biasanya dipilih $K = 5$ atau $k = 10$.
2. Untuk setiap $k = 1, 2, \dots, K$.
 - a. Data *training* merupakan data pada $(X_i, Y_i), i \neq t_k$ dan data validasi merupakan data pada $(X_i, Y_i), i = t_k$.
 - b. Untuk setiap nilai dari parameter *tuning* $\lambda \in \{\lambda_1, \dots, \lambda_m\}$ cari estimasi model $\hat{f}_\lambda$ menggunakan data training, yaitu model dengan tidak memasukkan data T yang terdapat pada *fold* ke- k .
 - c. Untuk setiap parameter λ hitung nilai $\hat{f}(x)$ untuk observasi pada data validasi atau data pada *fold* ke- k dan hitung *error* totalnya:

$$e_k(\lambda) = \sum_{i \in T_k} (y_i - \hat{f}_\lambda(x_i))^2$$

Keterangan:

Y_i : variabel respon ke- i

$\hat{f}_\lambda(x)$: nilai prediksi y untuk x_i pada saat *fold* ke- k dengan parameter λ

3. Untuk setiap parameter *tuning* λ , dihitung rata-rata *error* untuk semua *fold* adalah sebagai berikut:

$$CV(\lambda) = \frac{1}{n} \sum_{k=1}^K e_k(\lambda) = \frac{1}{n} \sum_{k=1}^K \sum_{i \in T_k} (Y_i - \hat{f}_\lambda(X_i))^2 \quad (23)$$

dengan n adalah banyaknya observasi.

1.4.10 Pengujian Signifikansi

Pengujian signifikansi parameter dilakukan untuk mengetahui variabel prediktor yang berpengaruh secara signifikan terhadap variabel respon (L. N. Azizah, 2013). Uji t merupakan suatu pengujian signifikansi untuk mengetahui pengaruh variabel prediktor terhadap variabel respon. Adapun hipotesis untuk uji t sebagai berikut.

$$H_0: \beta_j = 0$$

$$H_1: \beta_j \neq 0$$

untuk $j = 1, 2, \dots, k$

$$t_{hitung} = \frac{\hat{\beta}_j}{Se(\hat{\beta}_j)} \quad (24)$$

dengan

$$Se(\hat{\beta}_j) = \sqrt{var \hat{\beta}_j}$$

Kriteria ujinya yaitu, H_0 ditolak jika $|t_{hitung}| > t_{tabel}$ dengan k adalah banyak parameter atau $p - value < \alpha$ (Deria et al., 2019).

1.4.11 Uji Kelayakan Model

Metode yang digunakan dalam uji kelayakan model menggunakan nilai *Mean Square Error* (MSE) atau rerata jumlahan kuadrat *error* merupakan ukuran yang menunjukkan *error* dari nilai prediksi model terhadap dengan nilai observasi, atau seberapa jauh nilai prediksi menyimpang dari nilai yang sebenarnya. Nilai MSE dapat dicari melalui rumus berikut:

$$MSE = \frac{SSE}{n} = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \quad (25)$$

Dimana n merupakan banyaknya observasi, Y_i menunjukkan nilai observasi ke- i dan $\hat{Y}_i$ menunjukkan nilai prediksi ke- i . nilai MSE ini seringkali digunakan untuk menentukan model mana yang menjadi model terbaik dimana semakin kecil nilai MSE-nya maka model tersebut semakin baik. Sehingga yang merupakan model terbaik adalah model yang memiliki nilai MSE terkecil.

1.4.12 Indeks Pembangunan Manusia

Indeks Pembangunan Manusia (IPM) merupakan indikator yang digunakan untuk mengukur kualitas hidup masyarakat di suatu wilayah berdasarkan tiga dimensi utama, yaitu kesehatan, pendidikan, dan standar hidup layak. Di Indonesia, IPM menjadi tolok ukur dalam menilai keberhasilan Pembangunan. Nilai yang terus meningkat dari tahun ke tahun dengan berbagai kebijakan yang mendukung peningkatan kualitas pendidikan, kesehatan, dan kesejahteraan ekonomi. Dimensi

kesehatan dalam IPM diwakili oleh umur harapan hidup (UHH), yang mencerminkan rata-rata perkiraan tahun hidup seseorang sejak lahir berdasarkan kondisi kesehatan dan layanan medis yang tersedia. Dimensi pendidikan diukur melalui rata-rata lama sekolah (RLS), yaitu jumlah tahun rata-rata yang telah ditempuh oleh penduduk berusia 15 tahun ke atas dalam mengikuti pendidikan formal, serta harapan lama sekolah (HLS), yang menunjukkan jumlah tahun pendidikan yang diharapkan akan ditempuh oleh anak usia sekolah. Sementara itu, dimensi standar hidup diukur menggunakan pengeluaran per kapita disesuaikan, yang mencerminkan daya beli masyarakat dalam memenuhi kebutuhan dasar seperti pangan, sandang, dan papan. Selain ketiga dimensi utama tersebut, faktor eksternal seperti tingkat pengangguran terbuka (TPT) juga berpengaruh terhadap pembangunan manusia, karena mencerminkan ketersediaan lapangan kerja serta tingkat kesejahteraan ekonomi masyarakat.

Di Provinsi Sulawesi Selatan, perkembangan IPM menunjukkan peningkatan, meskipun masih terdapat perbedaan antar daerah. Kabupaten/kota dengan akses pendidikan dan fasilitas kesehatan yang lebih baik cenderung memiliki IPM yang lebih tinggi. RLS dan Pengeluaran per Kapita Disesuaikan menjadi faktor krusial dalam meningkatkan IPM, karena akses pendidikan yang lebih lama dapat meningkatkan keterampilan tenaga kerja dan berkontribusi terhadap peningkatan pendapatan masyarakat. UHH di Sulawesi Selatan juga mengalami perbaikan seiring dengan peningkatan layanan kesehatan. Namun, TPT di beberapa daerah masih menjadi tantangan akibat keterbatasan peluang kerja, terutama di daerah pedesaan. Dengan demikian, upaya peningkatan IPM di Sulawesi Selatan perlu diarahkan pada pemerataan akses pendidikan, peningkatan layanan kesehatan, serta penguatan sektor ekonomi guna menciptakan lebih banyak lapangan kerja yang berkelanjutan.

BAB II METODE PENELITIAN

2.1 Sumber data

Data yang digunakan dalam penelitian ini adalah data sekunder yang merupakan data kuantitatif mengenai Indeks Pembangunan Manusia (IPM) di Provinsi Sulawesi Selatan pada Tahun 2021. Data tersebut diperoleh dari situs resmi Badan Pusat Statistika (BPS) Provinsi Sulawesi Selatan tahun 2021 (<https://data.bps Sulsel.id/>). Unit penelitian yang digunakan dalam penelitian ini adalah 24 kabupaten/kota Provinsi Sulawesi Selatan yang terdiri dari 21 Kabupaten dan 3 Kota.

2.2 Variabel penelitian

Dalam penelitian ini, variabel yang digunakan terdiri dari 1 variabel respon (Y) dan 5 variabel prediktor (X) dengan rincian sebagai berikut:

Tabel 1. Variabel Penelitian.

Variabel	Nama Variabel	Definisi Operasional	Satuan
Y	Indeks Pembangunan Manusia	Indeks komposit yang mengukur pencapaian pembangunan manusia berdasarkan dimensi kesehatan, pendidikan, dan standar hidup.	Skala indeks (0-100)
X_1	Rata-rata Lama Sekolah (Tahun)	Rata-rata jumlah tahun yang telah ditempuh oleh penduduk usia 15 tahun ke atas dalam menempuh pendidikan formal.	Tahun
X_2	Pengeluaran per Kapita Disesuaikan (Ribu Rupiah/Orang/Tahun)	Rata-rata pengeluaran masyarakat per orang per tahun yang telah disesuaikan dengan daya beli.	Ribu Rupiah/Orang/Tahun
X_3	Umur Harapan Hidup (Tahun)	Rata-rata perkiraan umur yang dapat dicapai oleh seorang bayi yang lahir saat ini berdasarkan pola kematian yang ada.	Tahun
X_4	Harapan Lama Sekolah	Perkiraan jumlah tahun yang diharapkan dapat ditempuh oleh anak usia sekolah dalam jenjang pendidikan formal.	Tahun
X_5	Tingkat Pengangguran Terbuka	Persentase jumlah angkatan kerja yang tidak bekerja tetapi sedang mencari pekerjaan terhadap total angkatan kerja	Persentase (%)

2.3 Tahapan Analisis Data

Pada studi kasus ini proses analisis untuk data IPM di Provinsi Sulawesi Selatan tahun 2021 dilakukan dengan tahapan sebagai berikut:

1. Melakukan pengolahan data awal
 - Menentukan respon (Y) dan variabel prediktor (X).
 - Melakukan pengecekan multikolinearitas menggunakan VIF .
 - Standarisasi data jika variabel memiliki skala yang berbeda.
2. Mengestimasi awal dengan regresi OLS
 - Melakukan analisis regresi pada data dengan metode OLS untuk mendapatkan estimasi awal koefisien berdasarkan persamaan (7) yaitu :

$$\hat{\beta}_{OLS} = (X^T X)^{-1} X^T Y$$

estimasi digunakan untuk bobot dalam *adaptive* LASSO.

3. Melakukan perhitungan bobot (*weights*) untuk *adaptive* LASSO
 - Perhitungan bobot berdasarkan persamaan (20) yaitu:

$$W_j = \frac{1}{|\beta_j|^\gamma}$$

4. Melakukan Formulasi *adaptive* LASSO
 - Menerapkan regresi *adaptive* LASSO dengan penalti yang telah di beri bobot berdasarkan persamaan (19) yaitu:

$$\hat{\beta}_{ALASSO} = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^n \left(Y_i - \sum_{j=1}^k X_{ij} \beta_j \right)^2 + \lambda \sum_{j=1}^k W_j |\beta_j|$$

5. Melakukan implementasi *algoritma coordinate descent* analisis regresi pada data dengan metode *adaptive* LASSO.
6. Melakukan seleksi model dan evaluasi.
 - Menggunakan *Cross-Validation* (CV) untuk menentukan nilai optimal λ .
 - *K-fold Cross-Validation* digunakan untuk mencari λ yang meminimalkan error prediksi.
7. Mengecek multikolinearitas pada model hasil analisis dengan metode *adaptive* LASSO.
8. Melakukan uji signifikan dalam model.
9. Melakukan uji kebaikan model