

BAB I PENDAHULUAN

1.1 Latar Belakang

Analisis regresi merupakan salah satu metode analisis yang berfungsi untuk memodelkan hubungan antar variabel (Fernanda dkk., 2023). Dalam analisis regresi, variabel yang mempengaruhi disebut variabel prediktor (X), sedangkan variabel yang dipengaruhi disebut variabel respon (Y). Untuk mengevaluasi hubungan antara variabel prediktor dan variabel respon, teknik analisis yang digunakan adalah analisis regresi linear (Sarbanini dkk., 2022). Dalam model regresi linear, diasumsikan bahwa variabel respon mengikuti distribusi normal dan terdapat hubungan linear dengan variabel prediktor (Wigati, 2016).

Regresi linear merupakan metode pemodelan yang umum digunakan namun sering menghadapi kendala ketika asumsi-asumsi yang diperlukan tidak terpenuhi, seperti variabel respon diasumsikan tidak berdistribusi normal, atau terdapat hubungan non-linear antara variabel prediktor dan variabel respon (Ningsih, 2023). Dalam kondisi seperti ini, penting untuk mencari pendekatan alternatif yang dapat menghasilkan model yang valid dan akurat. Salah satu pengembangan dari model regresi linear yang digunakan untuk mengatasi keterbatasan tersebut adalah *Generalized Linear Model* (GLM), yang diperkenalkan oleh Nelder dan Wedderburn (1972). Model ini dapat menangani variabel respon yang tidak berdistribusi normal. Meski begitu, GLM tetap membutuhkan asumsi hubungan linear antara variabel prediktor dan variabel respon (Wahyuni, 2019). Namun dalam beberapa kasus penelitian, asumsi linearitas tidak selalu terpenuhi.

Masalah tersebut mendorong pengembangan GLM menjadi model yang lebih fleksibel, yang dikenal sebagai model aditif. Model tersebut dirancang untuk menangani data yang berdistribusi normal namun fungsi yang dimiliki tidak harus linear (Wigati, 2016). Model aditif kemudian dikembangkan menjadi model aditif tergeneralisasi (*Generalized Additive Model* (GAM)). Model tersebut lebih fleksibel karena mampu menangani hubungan non-linear dan variabel respon yang tidak harus berdistribusi normal (Wigati, 2016). Selain itu, GAM mampu memodelkan secara bersamaan efek linear dan non-linear dari berbagai variabel prediktor dalam satu kerangka model yang bersifat aditif dan fleksibel. Hal ini berbeda dengan metode *spline* yang hanya digunakan untuk memodelkan hubungan non-linear pada satu atau beberapa variabel secara terpisah. GAM memberikan kebebasan bagi setiap prediktor untuk dimodelkan dengan fungsi smoothing yang berbeda, sehingga dapat menangkap pola hubungan yang lebih kompleks serta interaksi antar variabel (Fithriasari dkk., 2018). Salah satu metode pendekatan yang dapat digunakan di dalam GAM adalah metode regresi nonparametrik.



parametrik merupakan metode pemodelan regresi yang digunakan atau pola data antara variabel respon dan prediktor tidak diketahui. Regresi parametrik, yang mana pola data linear, kuadrat, kubik, atau ditentukan sejak awal. Regresi nonparametrik bersifat fleksibel dan otomatis menyesuaikan dan menentukan bentuk kurva regresi. Oleh karena itu, diperlukan pendekatan khusus untuk mengestimasi nonparametrik (Islamiyati, 2014). Salah satu pendekatan estimator yang

sering digunakan dalam regresi nonparametrik adalah *penalized spline*. *Penalized spline* merupakan metode estimasi pada regresi *spline* yang diperoleh dengan meminimalkan fungsi *Penalized Least Square* (PLS). Metode ini dilengkapi dengan parameter *smoothing* yang berfungsi sebagai penghalus untuk mengatur keseimbangan antara kecocokan model terhadap data dan kelancaran kurva (Kurniasari dkk., 2019). Dengan adanya parameter *smoothing*, *penalized spline* mampu menyesuaikan pola data sekaligus mencegah terjadinya *overfitting* (Maulida, 2023).

Beberapa penelitian sebelumnya menunjukkan bahwa GAM telah diaplikasikan di berbagai bidang. Fernanda et al. (2023) menggunakan GAM untuk menganalisis data produksi kelapa sawit, Ningsih (2023) menerapkan GAM pada data investasi daerah, Valentina (2019) memanfaatkan GAM untuk memodelkan angka kematian bayi, Wahyuni (2016) menggunakan GAM dalam estimasi curah hujan, Biqytifa (2023) menerapkan GAM dalam pencadangan klaim asuransi, dan Latif (2022) menggunakan GAM untuk mengestimasi dan meramalkan jumlah kasus Covid-19. Selain bidang-bidang tersebut, GAM juga memiliki potensi besar untuk diterapkan dalam analisis Tingkat Pengangguran Terbuka (TPT), terutama untuk mengeksplorasi faktor-faktor yang mempengaruhinya.

Penelitian mengenai TPT sangat penting dilakukan, mengingat pengangguran adalah permasalahan sosial dan ekonomi yang dapat memengaruhi secara luas terhadap perkembangan dan keberlanjutan pembangunan nasional. TPT menjadi indikator penting untuk mengukur kondisi pasar tenaga kerja dan kesehatan perekonomian suatu negara, termasuk Indonesia, yang masih menghadapi tantangan pengangguran (Annam & Nasir, 2023). Pengangguran didefinisikan sebagai situasi ketika individu dalam angkatan kerja ingin bekerja tetapi belum berhasil mendapatkan pekerjaan (Adrian dkk., 2023). TPT merupakan indikator penting yang menggambarkan persentase pengangguran terhadap jumlah angkatan kerja (Agustina et al., 2022; Lestari et al., 2020). Berdasarkan Survei Angkatan Kerja Nasional (Sakernas) pada Agustus 2022, TPT Indonesia tercatat sebesar 5,86%, sementara Sulawesi Selatan berada di posisi ke-14 dari 34 provinsi dengan TPT sebesar 4,51%. Meskipun angka ini relatif rendah dibandingkan dengan provinsi lain, isu pengangguran tetap menjadi tantangan apabila dilihat dari perspektif kabupaten/kota. Dengan pemahaman yang lebih mendalam tentang faktor-faktor tersebut, diharapkan dapat menjadi referensi yang berguna dalam upaya mengurangi tingkat pengangguran di suatu daerah.

Penelitian mengenai TPT telah dilakukan oleh Zahro et al. (2021), menggunakan regresi linear untuk menganalisis faktor-faktor yang mempengaruhi TPT. Penelitian ini menunjukkan bahwa regresi linear mampu mengidentifikasi beberapa faktor yang berpengaruh, namun pendekatan tersebut memiliki keterbatasan dalam menangkap pola non-linear dan interaksi kompleks antar variabel. Oleh karena itu, diperlukan metode yang lebih fleksibel seperti GAM *spline* untuk mengeksplorasi faktor-faktor yang secara lebih komprehensif. Berdasarkan uraian tersebut, penelitian menerapkan GAM *spline* dalam memodelkan faktor-faktor yang di Provinsi Sulawesi Selatan. Dengan pendekatan ini, diharapkan wawasan yang lebih mendalam sebagai bahan rekomendasi dalam daerah tersebut.



1.2 Batasan Masalah

Batasan masalah pada penelitian ini adalah sebagai berikut:

1. Data yang digunakan adalah data tingkat pengangguran terbuka di Provinsi Sulawesi Selatan pada tahun 2022.
2. Estimator *spline* yang digunakan adalah *penalized spline*.
3. Penelitian ini menggunakan jumlah titik knot yang $k = 9$.

1.3 Tujuan dan Manfaat Penelitian

Tujuan dari penelitian ini adalah sebagai berikut:

1. Memperoleh estimasi parameter model *Generalized Additive Model* (GAM) dengan *penalized spline*.
2. Memperoleh model Tingkat Pengangguran Terbuka (TPT) di Sulawesi Selatan tahun 2022 dengan menggunakan metode *Generalized Additive Model* (GAM) dengan *penalized spline* beserta faktor-faktor yang mempengaruhinya.

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

1. Bagi peneliti, sebagai pengetahuan yang dapat memperluas pemahaman mengenai *Generalized Additive Model* (GAM) *spline* dan memberikan wawasan baru terkait penerapan metode tersebut pada kasus Tingkat Pengangguran Terbuka (TPT) di Provinsi Sulawesi Selatan.
2. Bagi pemerintah, sebagai acuan dalam mengevaluasi kebijakan serta mengambil langkah-langkah efektif untuk menurunkan Tingkat Pengangguran Terbuka (TPT) di Provinsi Sulawesi Selatan.
3. Bagi pembaca, sebagai referensi bagi pembaca yang tertarik dalam bidang statistika dan menjadi landasan bagi penelitian-penelitian selanjutnya yang relevan dengan topik ini.

1.4 Landasan Teori

1.4.1 Regresi Linear

Regresi linear merupakan metode statistik untuk membangun model yang menggambarkan hubungan antara satu variabel respon dengan satu atau lebih variabel prediktor (Fernanda dkk., 2023). Ketika hanya satu variabel prediktor digunakan dalam model, teknik ini dikenal sebagai regresi linear sederhana. Sedangkan jika model melibatkan beberapa variabel prediktor, maka teknik tersebut dikenal dengan regresi linear berganda (Alamsyah dkk., 2022). Model persamaan regresi linear berganda ditunjukkan pada Persamaan (1) (Ningsih, 2023):

$$y_i = \beta_0 + \sum_{t=1}^p \beta_t x_{it} + \varepsilon_i \quad (1)$$

$$i = 1, 2, 3, \dots, n; t = 1, 2, \dots, p$$



- n : banyaknya pengamatan
 p : banyaknya variabel prediktor.

1.4.2 Regresi Non-linear

Regresi non-linear merupakan metode statistik untuk memodelkan hubungan yang tidak mengikuti pola linear antara variabel respon dan variabel prediktor. Dalam konteks ini, regresi non-linear menggambarkan hubungan antara variabel prediktor dan variabel respon dengan cara yang mirip dengan persamaan diferensial, ketika persamaan tersebut berbentuk non-linear. Setiap model yang tidak linear dan memiliki parameter yang tidak diketahui dikategorikan sebagai regresi non-linear. Metode ini memungkinkan analisis hubungan yang lebih kompleks, memberikan fleksibilitas dalam menangkap pola data yang mungkin tidak dapat dijelaskan oleh model linear tradisional. Adapun persamaan model regresi non-linear dituliskan pada Persamaan (2) (Ningsih, 2023):

$$y_i = f(x_i\beta) + \varepsilon_i \quad (2)$$

dengan,

- y_i : nilai variabel respon pada pengamatan ke- i
 x_i : nilai variabel prediktor pada pengamatan ke- i
 β : parameter model regresi
 $f(x_i\beta)$: fungsi regresi dengan parameter β

1.4.3 Uji Linearitas

Uji linearitas merupakan suatu uji untuk menentukan apakah terdapat hubungan linear antara variabel prediktor dan variabel respon. Hubungan linear diartikan sebagai hubungan yang dapat digambarkan dengan garis lurus. Metode untuk memverifikasi hubungan linear dapat dilakukan uji *Ramsey's RESET*. Berikut hipotesis dari uji *Ramsey's RESET*:

H_0 : hubungan antara variabel prediktor dan respon adalah linear

H_1 : hubungan antara variabel prediktor dan respon adalah tidak linear

Adapun langkah – langkah dalam pengujian linearitas adalah sebagai berikut (Suhartono, 2008 dalam Ningsih, 2023):

- 1) Bentuk model regresi berdasarkan Persamaan (1) sehingga terbentuk pendugaan dari variabel respon seperti pada Persamaan (3):

$$\hat{y}_i = \beta_0 + \sum_{t=1}^p \beta_t x_{it} + \varepsilon_i \quad (3)$$

Kemudian hitung koefisien determinasi (R^2) dan notasikan dengan R_{old}^2

- 2) Bentuk model regresi berdasarkan Persamaan (1) dan tambahkan dua prediktor tambahan, yaitu $\hat{y}_i^2$ dan $\hat{y}_i^3$, sehingga menjadi Persamaan (4):

$$\hat{y}_i = \beta_0 + \sum_{k=1}^p \beta_k x_{ik} + \alpha_1 \hat{y}_i^2 + \alpha_2 \hat{y}_i^3 + \varepsilon_i \quad (4)$$

hitung koefisien determinasi (R^2) dan notasikan dengan R_{new}^2 .

statistik uji F menggunakan Persamaan (6):

$$F = \frac{\frac{R_{new}^2 - R_{old}^2}{m}}{\frac{1 - R_{new}^2}{n - p - m}} \quad (6)$$



dengan,

- R_{old}^2 : koefisien determinasi awal
 R_{new}^2 : koefisien determinasi akhir
 p : banyak prediktor awal
 m : banyak prediktor tambahan
 n : banyak pengamatan

Berikut kriteria pengambilan keputusannya (Widana & Muliani, 2020).

- Terima H_0 jika nilai $F < F_{\alpha; (df_1=m, df_2=n-p-1-m)}$ atau $p - value > 0,05$, maka terdapat hubungan yang linear antara kedua variabel.
- Tolak H_0 jika nilai $F > F_{\alpha; (df_1=m, df_2=n-p-1-m)}$ atau $p - value < 0,05$, maka tidak terdapat hubungan yang linear antara kedua variabel.

1.4.4 Generalized Linear Model (GLM)

Generalized Linear Model (GLM) merupakan pengembangan dari model regresi linear yang mengandung komponen acak, komponen sistematis, dan fungsi penghubung. GLM bertujuan untuk mengetahui hubungan sebab-akibat atau pengaruh dari variabel prediktor terhadap variabel respon. Keunggulan GLM dibandingkan dengan regresi linear biasa terletak pada distribusi (bentuk kurva) variabel respon. Variabel respon pada GLM tidak diasumsikan berdistribusi normal, melainkan distribusi-distribusi yang termasuk keluarga eksponensial (Sefle dkk., 2014). GLM terdiri dari 3 komponen utama sebagai berikut (Zahro et al., 2018):

- Komponen sistem linear (*linear predictor*)

GLM memiliki prediktor linear yang dituliskan pada Persamaan (7):

$$\eta_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} = x_i^T \beta \quad (7)$$

dengan,

η_i : prediktor linear untuk pengamatan ke- i

x_i : vektor prediktor ke- i

β : parameter model

- Komponen distribusi (keluarga eksponensial)

Variabel respon y_i dengan rata-rata $\mu_i = E(y_i)$, diasumsikan saling bebas dan berasal dari keluarga distribusi eksponensial yang meliputi distribusi normal, poisson, binomial, gamma, atau invers gaussian.

- Komponen fungsi *link*

Hubungan antara nilai harapan μ_i dan variabel prediktor linear η_i dinyatakan melalui sebuah fungsi link diferensiabel yang dituliskan pada Persamaan (8):

$$g(\mu_i) = \eta_i \quad (8)$$

dengan,

g : fungsi penhubung khusus yang menggambarkan distribusi dari variabel



Additive Model (GAM)

Merupakan pendekatan pemodelan yang menawarkan fleksibilitas lebih dari *Generalized Linear Model* (GLM). Pada GLM, hubungan antara prediktor diharuskan berbentuk linear, sehingga model tersebut

cenderung kaku dalam menangkap pola hubungan yang kompleks. Sebaliknya, dalam *Additive Models* setiap prediktor dihubungkan dengan respon melalui fungsi penghalus atau fungsi nonparametrik, sehingga hubungan tersebut dapat berbentuk non-linear sesuai dengan struktur data yang sebenarnya.

Model linear dan model aditif dapat digunakan untuk menganalisis berbagai jenis data. Namun, keduanya memiliki keterbatasan, salah satunya adalah asumsi bahwa data mengikuti distribusi normal. Permasalahan muncul ketika data yang dianalisis bersifat diskrit, karena distribusi normal tidak lagi sesuai. Oleh karena itu, dibutuhkan model yang lebih fleksibel dan mampu menangani distribusi data yang lebih beragam, yaitu *Generalized Additive Model (GAM)* (Wahyuni, 2016).

GAM merupakan pengembangan dari GLM dengan menggantikan komponen linear pada prediktor dengan jumlah fungsi penghalus yang fleksibel. Dengan demikian, GAM menyederhanakan regresi linear dengan menghapus asumsi mengenai bentuk fungsi yang digunakan untuk menggambarkan hubungan antara variabel respon dan prediktor (Fernanda dkk., 2023). Bentuk umum dari GAM dinyatakan pada Persamaan (9) (Wood, 2017):

$$\begin{aligned} y_i &= \beta_0 + f_1(x_1) + f_2(x_2) + \dots + f_p(x_p) + \varepsilon \\ &= \beta_0 + \sum_{t=1}^p f_t(x_t) + \varepsilon \end{aligned} \quad (9)$$

dengan,

y_i : nilai variabel respon ke- i

$f_t(x_t)$: fungsi penghalus (*smoothing function*) prediktor ke- t

ε : galat

1.4.6 Regresi Nonparametrik Spline

Regresi nonparametrik merupakan metode regresi yang digunakan untuk memodelkan hubungan antara variabel respon dan variabel prediktor ketika bentuk kurva regresinya belum diketahui. Metode ini memiliki tingkat fleksibilitas yang tinggi karena kurva regresi diasumsikan berada dalam suatu ruang fungsi tertentu, sehingga data akan membentuk sendiri pola kurva yang paling sesuai melalui teknik *smoothing* (penghalusan). Salah satu keunggulan regresi nonparametrik adalah tidak bergantung pada asumsi klasik sebagaimana pada regresi parametrik. Model umum regresi nonparametrik dapat dinyatakan pada Persamaan (10) (Härdle, 1994):

$$y_i = f(x_i) + \varepsilon_i, i = 1, 2, \dots, n \quad (10)$$

dengan

$f(x_i)$: fungsi regresi dengan bentuk kurva regresi yang tidak diketahui

Adapun bentuk umum regresi nonparametrik untuk kasus dengan lebih dari satu variabel prediktor, yang dikenal sebagai regresi nonparametrik multivariat, dapat dinyatakan pada Persamaan (11) (Dani et al., 2021):

$$y_i = \sum_{t=1}^p f_t(x_{it}) + \varepsilon_i \quad (11)$$



$f_t(x_{it})$: fungsi regresi nonparametrik prediktor ke- t pada pengamatan ke- i

ε_i : error regresi pada pengamatan ke- i

i : banyak pengamatan, untuk $i = 1, 2, \dots, n$

t : banyak variabel prediktor, untuk $t = 1, 2, \dots, p$

Estimasi fungsi pada regresi nonparametrik dapat menyesuaikan pola data yang sebenarnya dengan menggunakan metode *smoothing* (penghalus). Salah satu metode regresi nonparametrik yang memiliki fleksibilitas tinggi dalam memodelkan hubungan yang tidak diketahui adalah *spline*. Metode ini menggunakan segmen-segmen polinomial yang disambungkan secara kontinu melalui titik-titik knot untuk merepresentasikan data. Dengan keberadaan titik knot, *spline* mampu menyesuaikan bentuk kurva terhadap berbagai pola data, sehingga efektif dalam menangkap hubungan yang kompleks dan belum teridentifikasi sebelumnya. Titik knot berfungsi sebagai acuan dalam pembentukan serta penyesuaian kurva *spline*, membantu mengatasi pola data yang fluktuatif atau tidak beraturan, dan memberikan fleksibilitas lebih tinggi dibandingkan polinomial biasa (Ferryan, 2023). Dalam regresi *spline*, salah satu bentuk basis fungsi yang dapat digunakan adalah basis *truncated*. Basis *truncated* merupakan jenis basis fungsi *spline* yang berarti potongan fungsi (Fadhilah dkk., 2016). Secara umum, model fungsi regresi *spline* dapat dinyatakan pada Persamaan (12) (Kurniasari dkk., 2019):

$$f(x_i) = \beta_0 + \sum_{h=1}^r \beta_h x_i^h + \sum_{m=1}^k \beta_{(r+m)} (x_i - k_m)_+^r \quad (12)$$

dengan fungsi *truncated* yang diberikan:

$$(x_i - k_m)_+^r = \begin{cases} (x_i - k_m)^r, & x_i > k_m \\ 0, & x_i \leq k_m \end{cases}$$

Persamaan regresi nonparametrik spline diperoleh dengan menyubstitusikan Persamaan (12) ke dalam Persamaan (10), sehingga menghasilkan bentuk Persamaan (13):

$$y_i = \sum_{h=1}^r \beta_h x_i^h + \sum_{m=1}^k \beta_{(r+m)} (x_i - k_m)_+^r + \varepsilon_i \quad (13)$$

dengan,

k_m : titik knot ke- m

m : banyak titik knot, untuk $m = 1, 2, \dots, k$

h : orde, untuk $h = 1, 2, \dots, r$

$(x_i - k_m)_+^r$: fungsi *truncated* variabel ke- i dengan letak titik knot k_m dan orde r

Salah satu permasalahan yang sering terjadi dalam regresi *spline* adalah *overfitting*, yaitu kondisi ketika model terlalu menyesuaikan diri dengan data yang ada. Masalah ini biasanya muncul jika jumlah titik knot yang digunakan terlalu banyak, sehingga model menjadi sangat rumit. Meskipun hasil estimasi terhadap data pelatihan pun prediksi terhadap data baru menjadi rendah, terutama saat iasi yang tidak terdapat dalam data pelatihan. Untuk mengatasi hal $\Rightarrow$ *penalized spline*.



spline (P-spline) merupakan pendekatan statistik yang sibilitas model nonparametrik dengan kemampuan generalisasi bekerja dengan menyesuaikan kurva secara optimal terhadap data kat kehalusan kurva (Ferryan, 2023). Kurva *penalized spline*

diperoleh dengan meminimalkan fungsi *Penalized Least Square* (PLS), yaitu kriteria estimasi yang mengombinasikan antara kecocokan model terhadap data dan penalti untuk mengontrol kompleksitas model. Fungsi penalti ini dilengkapi dengan parameter penghalus (λ) yang berperan dalam menghindari *overfitting* (Eldeeb dkk., 2021). Adapun fungsi PLS dituliskan pada Persamaan (14):

$$PLS = \sum_{i=1}^n (y_i - f(x_i))^2 + \lambda \sum_{m=1}^k \beta_{r+m}^2 \quad (14)$$

dengan λ adalah parameter penghalus yang digunakan untuk mengatur tingkat kelengkungan kurva asli estimasi, dengan nilai $\lambda \geq 0$. Sebelum melakukan pencarian nilai optimal, fungsi PLS pada Persamaan (14) dapat ditulis ulang dalam bentuk matriks yang dituliskan Persamaan (15):

$$\begin{aligned} PLS &= (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) + \lambda \boldsymbol{\beta}^T \mathbf{D} \boldsymbol{\beta} \\ &= \mathbf{Y}^T \mathbf{Y} - 2\boldsymbol{\beta}^T \mathbf{X}^T \mathbf{Y} + \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{X} \boldsymbol{\beta} + \lambda \boldsymbol{\beta}^T \mathbf{D} \boldsymbol{\beta} \end{aligned} \quad (15)$$

dengan $\mathbf{D}$ merupakan matriks diagonal yang didefinisikan sebagai berikut:

$$\mathbf{D} = \begin{pmatrix} 0_{(r+1) \times (r+1)} & 0_{(r+1) \times k} \\ 0_{k \times (r+1)} & I_{k \times k} \end{pmatrix} = \begin{pmatrix} \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{I} \end{pmatrix}$$

dengan $\mathbf{I}$ merupakan matriks identitas untuk $\beta_{r+1}, \beta_{r+2}, \dots, \beta_{r+m}$.

Estimasi parameter regresi $\boldsymbol{\beta}$ diperoleh dengan melakukan diferensiasi terhadap fungsi PLS pada Persamaan (15) terhadap $\boldsymbol{\beta}^T$ sebagai berikut:

$$\begin{aligned} \frac{\partial PLS}{\partial \boldsymbol{\beta}^T} &= 0 \\ \frac{\partial (\mathbf{Y}^T \mathbf{Y} - 2\boldsymbol{\beta}^T \mathbf{X}^T \mathbf{Y} + \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{X} \boldsymbol{\beta} + \lambda \boldsymbol{\beta}^T \mathbf{D} \boldsymbol{\beta})}{\partial \boldsymbol{\beta}^T} &= 0 \\ 0 - 2\mathbf{X}^T \mathbf{Y} + 2\mathbf{X}^T \mathbf{X} \boldsymbol{\beta} + 2\lambda \mathbf{D} \boldsymbol{\beta} &= 0 \\ \hat{\boldsymbol{\beta}} &= (\mathbf{X}^T \mathbf{X} + \lambda \mathbf{D})^{-1} \mathbf{X}^T \mathbf{Y} \end{aligned} \quad (16)$$

Sehingga diperoleh estimasi model regresi yang dinyatakan pada Persamaan (17):

$$\hat{\mathbf{y}} = \mathbf{X} \hat{\boldsymbol{\beta}} = \mathbf{X} (\mathbf{X}^T \mathbf{X} + \lambda \mathbf{D})^{-1} \mathbf{X}^T \mathbf{Y} \quad (17)$$

1.4.7 Titik Knot dan Parameter Penghalus Optimal

Dalam regresi *penalized spline*, penentuan posisi titik knot didasarkan pada kuantil dari nilai-nilai unik variabel prediktor. Ketika digunakan sebanyak k titik knot, maka kuantil-kuantil tersebut akan membagi data pengamatan menjadi $k + 1$ bagian yang sama besar. Titik-titik knot ditentukan berdasarkan kuantil ke- $\frac{m}{k+1}$ dari himpunan nilai unik prediktor x_i , dengan $m = 1, 2, \dots, k$ (Nurfadhilah, 2023).

Parameter penghalus (λ) merupakan parameter pengontrol keseimbangan antara kemulusan fungsi dan kesesuaian fungsi terhadap data. Nilai λ yang semakin silkan estimasi fungsi yang lebih halus dan kurang responsif a, sedangkan nilai λ yang kecil memungkinkan fungsi menjadi lebih uti variasi pengamatan secara lebih ketat (Maulida, 2023). Oleh penting untuk menentukan λ pada tingkat yang optimal agar fungsi ang antara kelancaran dan akurasi.



Salah satu cara untuk menentukan jumlah titik knot dan parameter penghalus adalah melalui *Generalized Cross-Validation* (GCV), yang dinyatakan pada Persamaan (18) (Nurfadhilah, 2023):

$$GCV = \frac{MSE}{\left(1 - \frac{1}{n} \text{tr}[S(\lambda)]\right)^2} = \frac{n^{-1} \sum_{i=1}^n (y_i - \hat{y}_i)^2}{\left(1 - \frac{1}{n} \text{tr}[S(\lambda)]\right)^2} \quad (18)$$

dengan

$$\text{tr}[S(\lambda)] = \text{tr}[X(X^T X + n\lambda D)^{-1} X^T] \quad (19)$$

dengan,

MSE : Mean Square Error

λ : parameter penghalus (*smoothing parameter*)

y_i : variabel respon pada pengamatan ke- i

$\hat{y}_i$: nilai estimasi variabel respon pada pengamatan ke- i

i : banyak pengamatan, untuk $i = 1, 2, \dots, n$

$S(\lambda)$: matriks hat yang memuat parameter penghalus

D : matriks yang tidak menghasilkan nilai negatif, dan bagian diagonalnya diatur sesuai banyaknya titik knot pada model

1.4.8 Generalized Additive Model dengan Penalized Spline

Model *Generalized Additive Model* (GAM) dengan *penalized spline* merupakan pendekatan yang menggabungkan fleksibilitas *spline* dengan penalti untuk mengontrol kompleksitas model. Pendekatan ini memungkinkan untuk menangkap pola non-linear dalam data serta mengurangi risiko *overfitting*. Dalam konteks ini, model dinyatakan dalam Persamaan (9) dengan $f_t(x_t)$ merupakan fungsi *penalized spline* yang menggambarkan hubungan non-linear

Spline merupakan suatu fungsi yang dibentuk dari gabungan beberapa fungsi yang lebih sederhana. Fungsi-fungsi sederhana ini disebut sebagai fungsi basis, dan keseluruhannya dikenal sebagai basis. Dalam pemodelan *spline*, setiap fungsi basis b_t dikalikan dengan koefisien β_t , sehingga *spline* yang terbentuk merupakan hasil penjumlahan dari fungsi basis yang diberi bobot dan dievaluasi pada nilai x . Secara matematis, *spline* dinyatakan dalam Persamaan (20):

$$f(x) = \sum_{t=1}^B \beta_t b_t(x_t) \quad (20)$$

dengan,

B : jumlah basis *functions*

$b_t(x_t)$: fungsi basis *spline*

β_t : koefisien yang terkait dengan basis *spline*

Penalti (λ) diterapkan pada koefisien *spline* untuk mengontrol kelengkungan penalti derajat kedua yang dinyatakan pada Persamaan (21):

$$P(f_t) = \lambda_t \int \left(f_t''(x_t)\right)^2 dx_t \quad (21)$$



Penalti yang mengatur kekuatan penalti

atau dari fungsi *penalized spline* yang mengukur kelengkungan

Estimasi parameter dalam GAM dengan *penalized spline* dilakukan dengan meminimalkan fungsi objektif yang menggabungkan *Residual Sum of Squares* (RSS) dan penalti. Fungsi objektif dituliskan pada Persamaan (22):

$$RSS(\beta) = (Y - X\beta)^T(Y - X\beta) + \sum_{t=1}^p P(f_t) \quad (22)$$

dengan,

Y : vektor pengamatan

X : matriks desain yang mencakup variabel prediktor dan komponen *spline*

β : vektor parameter

Dengan pendekatan ini, model GAM dengan *penalized spline* memberikan keunggulan dalam menangkap pola non-linear yang kompleks sambil tetap menjaga stabilitas model. Pendekatan ini sangat berguna dalam analisis data yang memiliki variasi ekstrim atau pola non-linear yang tidak dapat diakomodasi oleh model regresi linear tradisional. Selain itu, penggunaan penalti dalam model ini membantu memastikan bahwa fungsi yang dihasilkan tidak terlalu rumit, sehingga meningkatkan kemampuan generalisasi model terhadap data baru.

1.4.9 Uji Signifikansi Parameter

Pengujian signifikansi parameter bertujuan untuk mengidentifikasi hubungan antara variabel respon dengan variabel prediktor, apakah berpengaruh signifikan atau tidak. Terdapat dua jenis uji signifikansi yang umum digunakan yakni uji simultan dan uji parsial.

1. Uji Simultan

Pengujian signifikansi parameter secara simultan atau yang biasa dikenal dengan uji F merupakan pengujian parameter model regresi secara bersamaan. Uji ini bertujuan untuk mengetahui parameter model regresi tersebut telah signifikan atau tidak (Polanagau dkk., 2019). Pengujian uji signifikansi simultan dapat dilakukan dengan hipotesis yaitu:

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_j = 0$$

$$H_1 : \text{minimal terdapat satu } \beta_j \neq 0, j = 1, 2, \dots, r + k$$

Adapun statistik uji yang digunakan dituliskan pada Persamaan (23) (Wongkar dkk., 2023):

$$F_{hitung} = \frac{MS_{Regresi}}{MS_{Error}} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2 / (r + k)}{\sum_{i=1}^n (y_i - \hat{y}_i)^2 / (n - (r + k) - 1)} \quad (23)$$

Tolak H_0 jika $F_{hitung} \geq F_{\alpha, (r+k, n-(r+k)-1)}$ atau $p\text{-value} < \alpha$.

2. Uji Parsial



ansi parameter secara parsial bertujuan untuk mengetahui apakah prediktor berpengaruh secara parsial terhadap variabel respon (2019). Pengujian uji signifikansi parsial dapat dilakukan dengan

Adapun statistik uji yang digunakan dinyatakan pada Persamaan (24) (Wongkar dkk., 2023):

$$t_{hitung} = \frac{\hat{\beta}_k}{SE(\hat{\beta}_k)} \quad (24)$$

Tolak H_0 jika $|t_{hitung}| > t_{\frac{\alpha}{2}, (n-(r+k)-1)}$ atau $p - value < \alpha$.

1.4.10 Tingkat Pengangguran Terbuka

Pengangguran (*unemployment*) merupakan keadaan individu yang termasuk dalam kategori angkatan kerja yang secara aktif mencari pekerjaan, tetapi belum berhasil mendapatkan pekerjaan (Sukirno, 2016 dalam Rahman & Riani, 2023). Indikator yang dapat digunakan untuk mengukur tingkat pengangguran disebut Tingkat Pengangguran Terbuka (TPT) (Agustina dkk., 2022). Menurut Badan Pusat Statistik (BPS), TPT merupakan proporsi orang yang menganggur terhadap keseluruhan angkatan kerja. TPT yang tinggi menunjukkan bahwa sebagian besar angkatan kerja belum terlibat dalam pasar kerja (Annam & Nasir, 2023). TPT berfungsi sebagai indikator kesehatan ekonomi dan pasar kerja, membantu pemerintah merumuskan kebijakan, mengevaluasi kinerja pasar, serta mengidentifikasi masalah pengangguran potensi pasar tenaga kerja di berbagai wilayah. Beberapa faktor yang diduga mempengaruhi TPT berdasarkan penelitian-penelitian terdahulu adalah:

1. Persentase Penduduk Miskin: Persentase penduduk yang tidak memiliki kemampuan untuk memenuhi standar minimum kebutuhan dasar (Lestari dkk., 2023).
2. Tingkat Partisipasi Angkatan Kerja (TPAK): Besarnya persentase penduduk dalam usia kerja yang ikut andil dengan aktif secara ekonomi di wilayah/daerah tertentu (Sari & Sugiharti, 2022).
3. Rata-rata Lama Sekolah (RLS): Durasi (dalam tahun) pendidikan yang telah dijalani oleh penduduk berusia 25 tahun ke atas (Sabrina dkk., 2022).
4. Produk Domestik Regional Bruto (PDRB): Jumlah nilai tambah yang diperoleh oleh semua unit usaha di suatu wilayah, atau keseluruhan nilai barang dan jasa akhir yang dihasilkan oleh seluruh unit ekonomi di suatu wilayah (Romhadhoni dkk., 2019).
5. Jumlah Penduduk: Jumlah orang yang bertempat tinggal atau menempati suatu desa/kota/sebagainya (Annam & Nasir, 2023).



BAB II METODE PENELITIAN

2.1 Sumber Data

Sumber data yang digunakan dalam penelitian ini merupakan data sekunder mengenai Tingkat Pengangguran Terbuka di Selatan pada tahun 2022, yang diperoleh dari Badan Pusat Statistik (BPS) Provinsi Sulawesi Selatan. Data tersebut dapat diakses di laman <https://sulsel.bps.go.id>. Unit observasi dalam penelitian ini adalah Kabupaten/Kota di Provinsi Sulawesi Selatan, yang terdiri dari 21 kabupaten dan 3 kota. Data lengkap dilampirkan pada Lampiran 1.

2.2 Variabel Penelitian

Variabel yang digunakan pada penelitian ini terdiri atas satu variabel respon dan lima variabel prediktor yang dituliskan pada Tabel 1.

Tabel 1. Variabel Penelitian

Variabel	Keterangan	Satuan	Definisi
Y	Tingkat Pengangguran Terbuka (TPT)	%	Persentase angkatan kerja yang tidak memiliki pekerjaan tetapi aktif mencari pekerjaan.
X_1	Persentase Penduduk Miskin (PPM)	%	Persentase jumlah penduduk yang hidup di bawah garis kemiskinan di suatu wilayah.
X_2	Tingkat Partisipasi Angkatan Kerja (TPAK)	%	Persentase penduduk usia kerja yang aktif bekerja atau mencari pekerjaan.
X_3	Rata-rata Lama Sekolah (RLS)	Tahun	Rata-rata lama pendidikan formal yang ditempuh oleh penduduk di suatu wilayah.
X_4	Produk Domestik Regional Bruto (PDRB)	Milyar Rupiah	Jumlah nilai tambah barang dan jasa akhir yang dihasilkan oleh seluruh unit ekonomi di wilayah tertentu dalam periode tertentu.
X_5	Jumlah Penduduk (JP)	Ribu Jiwa	Total jumlah penduduk yang tinggal di suatu wilayah.

2.3 Metode Analisis

Tahapan analisis data menggunakan *Generalized Additive Models (GAM) spline* berdasarkan tujuan penelitian terdiri dari beberapa langkah sebagai berikut:



1. Menentukan parameter model *Generalized Additive Model (GAM)* dengan menggunakan *spline*.

2. Melakukan analisis deskriptif terhadap Tingkat Pengangguran Terbuka (TPT) di Provinsi Sulawesi Selatan dengan menggunakan metode *Generalized Additive Model (GAM)* dengan *penalized spline* beserta faktor-faktor yang mempengaruhinya.

3. Melakukan analisis statistika deskriptif untuk menganalisis gambaran atau hubungan antara variabel respon dan variabel prediktor.

- b. Membuat *scatterplot* antara variabel respon dengan masing-masing variabel prediktor untuk mengetahui pola hubungan yang terjadi.
- c. Melakukan uji linearitas dengan menggunakan *Ramsey's Reset Test* pada Persamaan (5), untuk menentukan variabel prediktor yang bersifat linear dan non-linear terhadap variabel respon. Jika nilai Sig. > 0,05, maka terdapat hubungan yang linear antara kedua variabel dan jika nilai Sig. < 0,05, maka tidak terdapat hubungan yang linear antara kedua variabel.
- d. Menentukan titik knot dan parameter penghalus optimal berdasarkan nilai GCV terendah untuk mendapatkan kurva yang pas dan menghindari *overfitting* maupun *underfitting*.
- e. Melakukan pemodelan *Generalized Additive Model* (GAM) dengan menambahkan pemulus *penalized spline* pada variabel prediktor yang bersifat non-linear terhadap variabel respon.
- f. Melakukan pengujian signifikansi parameter secara simultan dan parsial.
- g. Membuat interpretasi hasil dan menarik kesimpulan.

