

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang Masalah

Metode dalam statistik yang populer digunakan salah satunya adalah analisis regresi. Analisis regresi biasanya digunakan untuk menentukan hubungan fungsional antara variabel independen (prediktor) dan variabel dependen (respon) (Rahmawati, D. P. et al., 2021). Model regresi terbagi ke dalam tiga bentuk, yaitu regresi parametrik, regresi semiparametrik, dan regresi nonparametrik. Regresi parametrik merupakan pendekatan model regresi, jika bentuk kurva atau fungsinya diketahui. Regresi nonparametrik merupakan pendekatan model regresi, jika bentuk kurva tidak diketahui atau pola datanya tidak diketahui (Risnawati, Fernandes, A. A. R., dan Nurjannah, 2019). Regresi semiparametrik merupakan regresi di mana sebagian bentuk pola diketahui dan sebagian lain bentuk pola data tidak diketahui, dengan kata lain model regresi semiparametrik merupakan penggabungan dari model regresi parametrik dan nonparametrik (Wahba, G., 1990).

Regresi parametrik memiliki asumsi yang kaku dan kuat, di mana bentuk kurva regresi diketahui, seperti bentuk kurva linear, kuadrat, dan bentuk kurva lainnya (Budiantara, I. N., 2009). Dalam regresi nonparametrik data akan mencari sendiri bentuk estimasi dari kurva regresinya dengan tidak dipengaruhi faktor subjektivitas peneliti, sehingga regresi nonparametrik fleksibel dan objektif (Budiantara, I. N., 2011). Pendekatan model regresi nonparametrik yang sering digunakan adalah regresi kernel, *wavelet*, *spline*, dan deret fourier (Risnawati, Fernandes, A. A. R., dan Nurjannah, 2019).

Model regresi nonparametrik yang banyak digunakan untuk melakukan penelitian oleh para peneliti salah satunya adalah regresi *spline*. *Spline* merupakan estimator yang memiliki fleksibilitas yang baik untuk mengestimasi fungsi regresi nonparametrik (Lestari, B. et al., 2018). *Spline* memiliki kemampuan yang sangat baik

dalam data yang perilakunya berubah-ubah pada subinterval tertentu (Risnawati, Fernandes, A. A. R., dan Nurjannah, 2019). Beberapa pendekatan regresi *spline*, di antaranya *B-Spline*, *naive spline*, *penalized spline*, *spline truncated*, dan lain-lain.

Dalam regresi *spline*, pemilihan letak titik *knot* dan jumlah titik *knot* merupakan hal yang penting karena akan berpengaruh terhadap keoptimalan model regresi (Wu, H., dan Zhang, J. T., 2006). Keoptimalan model dapat dipilih dengan menggunakan metode *Generalized Cross Validation* (GCV) yang merupakan pengembangan dari *Cross Validation* (CV), pemilihan keoptimalan model dilakukan dengan memilih GCV yang minimum. Dalam melakukan pemilihan letak titik *knot*, jika dilakukan secara manual akan dibutuhkan waktu yang lama. Untuk mengatasi hal tersebut dapat dilakukan menggunakan *penalized spline*, di mana dalam *penalized spline* titik *knot* berada di titik-titik kuantil dari nilai tunggal (*unique*) variabel prediktor (Ruppert, D., Wand, M. P., dan Carrol, R. J., 2003).

*Penalized spline* merupakan estimator dari regresi nonparametrik yang didapatkan dengan meminimumkan fungsi *Penalized Least Square* (PLS). *Penalized Least Square* (PLS) adalah metode optimasi yang memberikan komponen penghalus pada metode *least square* dengan kriteria optimasi yang menggabungkan antara kecocokan (*goodness of fit*) terhadap data dan kemulusan kurva (*penalty*) (Islamiyati, A., dan Herdiani, E. T., 2012). *Penalized spline* melibatkan titik *knot* dan parameter *smoothing* secara bersamaan untuk mengontrol kemulusan (*smoothness*) dari kurva. Keunggulan menggunakan *penalized spline* adalah kemampuan estimator untuk menghasilkan kurva yang *smooth* dan *fit* kurva (Islamiyati, A., Fatmawati, dan Chamidah, N., 2018).

Beberapa penelitian yang terkait dengan *penalized spline*, antara lain pada penelitian Widyowati, C. (2016) membahas tentang estimasi model regresi nonparametrik aditif berdasarkan estimator *penalized spline*, Andrianto, D. (2017) membahas estimasi model regresi semiparametrik dengan menggunakan *penalized spline*, Adiwati, T., dan Chamidah, N. (2019) membahas tentang pemodelan resiko hipertensi dengan mempertimbangkan faktor-faktor yang mempengaruhinya dengan

menggunakan regresi nonparametrik *penalized spline*, Putri, F. P. F. (2017) membahas estimasi model dengan menggunakan *penalized spline*.

Dalam analisis regresi, data yang digunakannya dibagi menjadi tiga jenis yaitu data *time series*, *cross section*, dan data longitudinal. Data *time series* (runtun waktu) merupakan jenis data yang dikumpulkan menurut urutan waktu dalam suatu rentang waktu tertentu. Data *cross section* merupakan jenis data yang didapat dengan mengamati banyak subyek dalam satu waktu yang sama. Data longitudinal adalah data yang terdiri dari data *cross-section* dan data *time series* (Diggle, P. J. et al., 2006). Regresi nonparametrik telah digunakan untuk menganalisis data *cross section*, seperti pada penelitian Nurdiani, N., Herrhyanto, N., dan Dasari, D. (2017) yang menggunakan estimator *spline* untuk mengetahui faktor-faktor yang mempengaruhi angka melek huruf dan rata-rata lama sekolah, Widyowati, C. (2016) melakukan pemodelan persentase kemiskinan di Jawa Timur tahun 2013 dengan menggunakan estimator *penalized spline*, Widyowati, C. (2016) mengestimasi kurva dengan estimator gabungan *spline truncated* dan deret fourier.

Dalam regresi *spline* data yang digunakan bisa berupa data respon tunggal, birespon, dan multirespon. Pada data respon tunggal bisa menggunakan prediktor tunggal atau multiprediktor. Pada data birespon dan multirespon bisa menggunakan data prediktor tunggal atau multiprediktor. Model birespon memiliki asumsi terdapat korelasi antar respon. Tujuan pemodelan birespon adalah untuk memperoleh model regresi yang lebih baik daripada model regresi tunggal, karena model birespon tidak hanya memperhatikan hubungan antara respon dan prediktor tapi juga memperhatikan hubungan antar respon (Femadiyanti, S. F., Suparti, dan Warsito, B., 2020). Penelitian tentang regresi nonparametrik birespon, antara lain Yolandika, B. C. (2011) yang melibatkan dua variabel respon dan satu variabel prediktor, penelitian Islamiyati, A., Fatmawati, Chamidah, N. (2020) yang melibatkan dua variabel respon dan tiga variabel prediktor, Putri, F. P. F. (2017) membahas estimasi model dengan menggunakan *penalized spline* dengan melibatkan dua variabel respon dan empat variabel prediktor. Pada penelitian sebelumnya, belum ada penelitian yang menggunakan dua variabel respon dan lima variabel prediktor padahal bisa saja terdapat lebih banyak prediktor

yang mempengaruhi variabel respon. Oleh karena itu, untuk pengembangan penelitian akan dilakukan penelitian dengan menggunakan *penalized spline* untuk data *cross section* dengan menggunakan dua variabel respon dan lima variabel prediktor.

Manusia merupakan kekayaan bangsa yang sesungguhnya sehingga tujuan akhir pembangunan harus difokuskan pembangunan manusia, di mana kondisi tersebut dapat menciptakan lingkungan yang memungkinkan masyarakat untuk dapat menikmati umur panjang, sehat, dan menjalankan kehidupan yang produktif (Badan Pusat Statistik, 2015). Indeks Pembangunan Manusia (IPM) menjelaskan bagaimana manusia bisa mengakses hasil pembangunan dalam memperoleh pendidikan, pendapatan, kesehatan. IPM dibentuk oleh tiga dimensi dasar yaitu standar hidup layak, umur panjang dan sehat, dan juga pengetahuan. IPM menjadi indikator untuk menilai aspek kualitas dari pembangunan dan mengklasifikasikan suatu negara termasuk negara maju, negara berkembang atau negara terbelakang dan untuk mengukur kebijakan ekonomi terhadap kualitas hidup (Utami, F. P., 2020).

Pada periode 1970-2010 Indonesia masuk ke dalam bagian *World's Top Movers in HDI improvement*, di mana pada periode tersebut Indonesia termasuk ke dalam negara yang mengalami kemajuan tercepat dalam peningkatan IPM. Pada 2020 Badan Pusat Statistik (BPS) mencatat adanya perlambatan pertumbuhan IPM Indonesia dibandingkan tahun sebelumnya dengan pertumbuhan sebanyak 0,03% atau kenaikan sebanyak 0,02 poin (Badan Pusat Statistik, 2020a). Berdasarkan data yang telah dianalisis oleh *United Nation Development Programme* (UNDP), IPM Indonesia tahun 2020 berada pada peringkat 107 dari 189 negara yang dianalisis dan berada pada peringkat 5 diantara negara-negara di ASEAN, di mana Indonesia berada di bawah Singapura, Brunei Darussalam, Malaysia dan Thailand. Dilihat dari indikator IPM, Indonesia tertinggal jauh oleh Singapura dari semua aspek. Jika dibandingkan dengan Brunei Darussalam dan Malaysia, Indonesia masih kalah pada semua aspek dari indikator IPM. Akan tetapi ketika dibandingkan dengan Thailand, tertinggal dalam aspek Angka Harapan Hidup (AHH) dan pengeluaran per kapita. Angka Harapan Hidup (AHH) merupakan indikator IPM yang menggambarkan umur panjang dan hidup sehat, sedangkan pengeluaran per kapita merupakan indikator IPM yang

menggambarkan standar hidup layak. Menurut Badan Pusat Statistik (2021b) angka harapan hidup merupakan rata-rata perkiraan banyak tahun yang dapat ditempuh oleh seseorang sejak lahir. Menurut Badan Pusat Statistik (2021a) pengeluaran per kapita adalah biaya yang dikeluarkan untuk konsumsi semua anggota rumah tangga selama sebulan dibagi dengan banyaknya anggota rumah tangga yang telah disesuaikan dengan paritas daya beli. Berdasarkan penjelasan tersebut, perlu dilakukan identifikasi faktor-faktor yang mempengaruhi angka harapan hidup dan pengeluaran per kapita, dengan tujuan faktor yang berpengaruh tersebut dapat lebih mendapatkan perhatian agar dapat meningkatkan IPM Indonesia seperti pada periode 1970-2010.

Berdasarkan paparan tersebut, dalam penelitian ini akan dibahas mengenai Estimasi Model Regresi Nonparametrik Birespon dengan *Penalized Spline* untuk Data *Cross Section*. Variabel respon yang akan dilibatkan adalah Angka Harapan Hidup (AHH) dan Pengeluaran Per Kapita Provinsi di Indonesia. Variabel prediktornya, yaitu Harapan Lama Sekolah, Rata-Rata Lama Sekolah, Laju Pertumbuhan Ekonomi, Tingkat Pengangguran Terbuka, Upah Minimum Provinsi. Data yang didapatkan merupakan data yang berasal dari Badan Pusat Statistik (BPS) dan Sistem Informasi dan Manajemen Data Dasar Regional (SIMREG) pada tahun 2020.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang masalah, maka permasalahan yang dibahas dalam penelitian ini adalah sebagai berikut:

1. Bagaimana model regresi nonparametrik birespon dengan *penalized spline* terhadap data angka harapan hidup dan pengeluaran per kapita?
2. Bagaimana analisa dan interpretasi model regresi nonparametrik birespon dengan *penalized spline* terhadap data angka harapan hidup dan pengeluaran per kapita?

### 1.3 Tujuan Penelitian

Berdasarkan perumusan masalah, maka tujuan yang ingin dicapai dalam penelitian ini adalah sebagai berikut:

1. Membuat model regresi nonparametrik birespon dengan *penalized spline* terhadap data angka harapan hidup dan pengeluaran per kapita
2. Menganalisa dan interpretasi model regresi nonparametrik birespon dengan *penalized spline* terhadap data angka harapan hidup dan pengeluaran per kapita.

### 1.4 Manfaat Penelitian

Berdasarkan tujuan penelitian ini, manfaat pada penelitian ini adalah sebagai berikut:

1. Manfaat Teoritis  
Menambah wawasan pengetahuan kepada penulis dan kepada masyarakat terkait model regresi nonparametrik birespon dengan *penalized spline* terhadap data angka harapan hidup dan pengeluaran per kapita
2. Manfaat Praktis  
Sebagai bahan masukan kepada Pemerintah Indonesia terkait dengan permasalahan angka harapan hidup dan pengeluaran per kapita

### 1.5 Batasan Masalah

Batasan masalah dalam penelitian ini adalah sebagai berikut:

1. Pada pemilihan jumlah titik *knot* menggunakan metode *fixed selection method*
2. Pemilihan model regresi nonparametrik optimal berdasarkan pada nilai GCV minimum.