

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Analisis regresi merupakan suatu metode yang digunakan untuk menganalisis hubungan fungsional antara variabel prediktor (X) dan variabel respon (Y), dimana teknik statistika ini telah digunakan secara luas dalam ilmu pengetahuan terapan. Hubungan fungsional kedua variabel tersebut dijelaskan dalam sebuah kurva yang dinamakan kurva regresi. Pendekatan yang digunakan dalam menentukan kurva regresi ada dua jenis, yaitu pendekatan statistika parametrik dan pendekatan statistika nonparametrik. Selain digunakan untuk mengetahui bentuk hubungan antar variabel, analisis regresi juga dapat digunakan untuk peramalan. (Hardle, 1990 : 4)

Pendekatan statistika parametrik digunakan jika asumsi sebaran data membentuk pola tertentu, misal berbentuk linear, kuadratik, kubik dan lain-lain. Asumsi tersebut didasarkan atas teori dan informasi yang spesifik dan kuantitatif dari peneliti mengenai bentuk fungsinya ataupun dari pengetahuan masa lalu. Jika asumsi pada kurva regresi dengan pendekatan statistika parametrik tidak dipenuhi, maka kurva regresi diduga menggunakan pendekatan statistika nonparametrik. Metode ini tidak bergantung pada asumsi-asumsi tertentu, seperti kenormalan suatu data, varians yang sama dan error yang tidak berkorelasi. Metode ini digunakan ketika informasi mengenai kurva regresi dari sekumpulan data sangat

sedikit atau bahkan tidak diketahui. Terdapat beberapa teknik untuk menaksir parameter dalam regresi nonparametrik, antara lain Histogram, estimator Kernel, estimator Deret Orthogonal dan estimator Spline.

Salah satu model regresi dengan pendekatan statistika nonparametrik yang dapat digunakan untuk menduga kurva regresi adalah regresi spline. Regresi spline adalah suatu pendekatan ke arah pencocokan data dengan tetap memperhitungkan kemulusan kurva. Spline adalah salah satu jenis *piecewise* polinomial, yaitu polinomial yang memiliki sifat tersegmen. Sifat tersegmen ini memberikan fleksibilitas lebih dari model polinomial biasa, sehingga memungkinkan untuk menyesuaikan diri secara lebih efektif terhadap karakteristik lokal suatu fungsi atau data. Dalam regresi spline, langkah awal yang dilakukan adalah menentukan knot dengan nilai *Generalized Cross Validation* (GCV) yang minimum. Selain melihat nilai GCV yang minimum, kriteria lain yang dapat digunakan adalah dengan melihat nilai *Mean Square Error* (MSE). (Lee, 2002)

Knot dapat diartikan sebagai suatu titik fokus dalam fungsi spline sehingga kurva yang dibentuk dapat terbagi pada titik tersebut. Dalam fungsi spline, memungkinkan digunakan berbagai macam orde, sehingga dapat dibentuk regresi spline linear, regresi spline kuadratik, dan seterusnya. Orde dalam fungsi spline diartikan sebagai pangkat terbesar dalam fungsi spline. Selain itu, dimungkinkan juga penggunaan knot sampai $n - 2$ buah titik sehingga dikenal regresi linear spline dengan 1 knot dan regresi kuadratik dengan 2 knot.

Bentuk umum fungsi spline orde ke- m adalah sebagai berikut :

$$f(x) = \beta_0 + \sum_{j=1}^m \beta_j X^j + \sum_{j=1}^k \lambda_j (X - K_j)_+^m$$

di mana :

β_0 = Konstanta

β_j = Koefisien pada variabel X^j

λ_j = Koefisien pada variabel X^j terbagi pada knot ke- j

X^j = Variabel prediktor orde ke- j

K_j = Knot ke- j pada variabel X^j

$j = 1, 2, \dots, m$ dan k = banyak knot.

Sedangkan bentuk umum regresi spline orde ke- m adalah

$$Y = \beta_0 + \sum_{j=1}^m \beta_j X^j + \sum_{j=1}^k \lambda_j (X - K_j)_+^m + \varepsilon$$

Spline mempunyai kelemahan pada saat orde spline tinggi. Ketika knot yang banyak dan knot yang terlalu dekat, hal itu akan membentuk matriks dalam perhitungan yang hampir singular, sehingga persamaan tidak dapat diselesaikan.

Penggunaan spline difokuskan pada pola data yang pada daerah tertentu mempunyai karakteristik yang berbeda dengan daerah yang lain. Dalam tugas akhir ini akan digunakan contoh data mengenai berat badan balita umur 0 bulan sampai 60 bulan pada bulan Mei 2011 yang diperoleh dari Posyandu Dewi Pramanik yang beralamat di Jl. Kopo Gg. Pabrik Kulit Utara RW01 Kelurahan Sukaasih Kecamatan Bojongloa Kaler Kota Bandung, kemudian akan dicari

model regresi spline menggunakan knot dan orde dengan melihat nilai GCV dan MSE yang minimum untuk nilai λ dengan menggunakan program pengolah data SAS 9.0. Oleh karena itu, tugas akhir ini diberi judul “Analisis Regresi Spline Menggunakan Knot dan Orde (Studi Kasus pada Data Berat Badan Balita di Posyandu Dewi Pramanik Kelurahan Sukaasih Kecamatan Bojongloa Kaler Kota Bandung)”.

1.2 Rumusan Masalah

Permasalahan yang akan dibahas dalam tugas akhir ini adalah :

1. Bagaimana cara regresi spline dalam menentukan kurva regresi?
2. Bagaimana model regresi spline terbaik yang memenuhi kriteria MSE (λ) dan GCV (λ)?
3. Bagaimana pengaruh lokasi knot terhadap MSE (λ) dan GCV (λ) pada model regresi spline?
4. Bagaimana penerapan regresi spline yang menggunakan knot dan orde dalam sebuah studi kasus mengenai berat badan balita umur 0 bulan sampai 60 bulan pada bulan Mei 2011 di Posyandu Dewi Pramanik RW 01 Kelurahan Sukaasih Kecamatan Bojongloa Kaler Kota Bandung?

1.3 Tujuan Penulisan

Berdasarkan rumusan masalah di atas, maka tujuan pembuatan tugas akhir ini adalah sebagai berikut :

1. Mengetahui cara regresi spline dalam menentukan kurva regresi.

2. Memilih model regresi spline terbaik yang memenuhi kriteria MSE (λ) dan GCV (λ).
3. Mengetahui pengaruh lokasi knot terhadap MSE (λ) dan GCV (λ) pada model regresi spline.
4. Mengetahui penerapan regresi spline yang menggunakan knot dan orde dalam sebuah studi kasus mengenai berat badan balita umur 0 bulan sampai 60 bulan pada bulan Mei 2011 di Posyandu Dewi Pramanik RW 01 Kelurahan Sukaasih Kecamatan Bojongloa Kaler Kota Bandung.

1.4 Batasan Masalah

Dalam pembahasan selanjutnya, masalah yang dibahas dalam tugas akhir ini dibatasi pada memodelkan dan memilih model regresi spline dengan satu variabel prediktor (X) berdasarkan kriteria MSE (λ) dan GCV (λ) yang terkecil. Untuk semua kemungkinan banyaknya knot dan penempatan knot akan dimodelkan regresi spline menggunakan fungsi spline dengan orde ke-2 (kuadratik) dan orde ke-3 (kubik).

Penaksiran parameter pada regresi spline dilakukan dengan cara meminimumkan persamaan *Penalized Least Square* (PLS). Dalam Tugas Akhir ini, penaksiran kurva regresi spline akan menggunakan program pengolah data SAS 9.0.

1.5 Manfaat Penulisan

1.5.1 Manfaat Teoritis

Pada kenyataannya, sekumpulan data tidak selalu berasal dari populasi berdistribusi normal dan bentuk kurvanya pun tidak diketahui sehingga pendekatan statistika parametrik tidak dapat digunakan. Oleh karena itu, pendekatan yang cocok digunakan dengan kriteria data tersebut adalah pendekatan statistika nonparametrik. Oleh karena itu, penelitian yang dilakukan diharapkan dapat meningkatkan pengetahuan tentang teori bentuk-bentuk penaksir pada regresi nonparametrik. Hasil dari penelitian juga diharapkan dapat menambah wawasan dan menjadi referensi akademik dari teori-teori yang dipelajari selama perkuliahan dan menjadi pelengkap informasi yang berguna bagi penelitian selanjutnya.

1.5.2 Manfaat Praktis

Melalui tugas akhir ini, akan diperoleh bentuk kurva regresi dari data pertumbuhan berat badan balita umur 0 bulan sampai 60 bulan pada bulan Mei 2011 di Posyandu Dewi Pramanik RW 01 Kelurahan Sukaasih Kecamatan Bojongloa Kaler Kota Bandung, yang kemudian akan diketahui model regresi spline menggunakan knot dan orde dengan kriteria MSE (λ) dan GCV (λ).