



**PEMODELAN REGRESI NONPARAMETRIK *SPLINE*
MULTIVARIABEL**

Skripsi

disusun sebagai salah satu syarat
untuk memperoleh gelar Sarjana Sains
Program Studi Matematika

oleh

Tiffani Dita Permata Putri

4111414009

**JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS NEGERI SEMARANG
2018**

PERNYATAAN

Saya menyatakan bahwa skripsi ini bebas plagiat, dan apabila di kemudian hari terbukti terdapat plagiat dalam skripsi ini, maka saya akan bersedia menerima sanksi sesuai ketentuan perundang-undangan.

Semarang, September 2018



Tiffani Dita Permata Putri
NIM 4111414009

PENGESAHAN

Skripsi yang berjudul

Pemodelan Regresi Nonparametrik *Spline* Multivariabel

disusun oleh

Tiffani Dita Permata Putri

4111414008

Telah dipertabangkan di hadapan sidang Panitia Ujian Skripsi FMIPA UNNES pada tanggal 12 September 2018.



Prof. Dr. Zaenuri, S.E., M.Si., Akt.
NIP. 196412231988031001

Sekretaris

Drs. Arief Agoestanto, M.Si.
NIP. 196807221993031005

Ketua Penguji

Dr. Scolastika Mariani, M.Si.
NIP 196502101991022001

Anggota Penguji/
Pembimbing I

Dr. Nur Karomah Dwidayati, M.Si.
NIP 196605041990202001

Anggota Penguji/
Pembimbing II

Dra. Sunarni, M.Si.
NIP 195506241988032001

MOTTO DAN PERSEMBAHAN

MOTTO

1. “Menuntut Ilmu itu wajib atas setiap muslim.” (HR. Ibnu Majah)
2. “*Break the limit.*” (GG Mild)
3. “Semangatlah, kamu tak akan tau apa yang akan terjadi esok.” (Penulis)

PERSEMBAHAN

1. Syukurku kupersembahkan kepada Allah SWT yang maha kuasa
2. Kedua orang tuaku Bapak Sugiyanto, Alm. Ibu Salminah, dan Ibu Tri Asih atas segala ridho, doa, dukungan, dan kasih sayang yang telah diberikan.
3. Untuk Adikku tersayang Rivo Restu Ananda Putra dan Kakakku tercinta Bagas Takum Triasmoro.
4. Mas Firstyan Ariful Rizal telah menyemangati hingga akhir.
5. Dosen-dosen jurusan Matematika dan dosen pembimbing yang sudah memberikan saya ilmu yang bermanfaat dalam menyelesaikan skripsi.
6. Keluarga besar Himpunan Mahasiswa Matematika
7. Untuk Retno Ayu, Leni, Nia Dwi Astuti, dan teman-teman Matematika Unnes 2014 terimakasih atas dukungan kalian.
8. Untuk almamaterku Universitas Negeri Semarang.

KATA PENGANTAR

Puji syukur kehadirat Allah SWT atas nikmat dan karunia-Nya Semoga Allah SWT selalu memberikan keridhoan di setiap jalan yang kita tempuh sehingga penulis dapat menyelesaikan skripsi dengan judul “Pemodelan Regresi Nonparametrik *Spline* Multivariabel. Penyusunan skripsi ini merupakan salah satu syarat untuk memperoleh gelar Sarjana Sains.

Dalam penyusunan skripsi ini tidak lepas dari bantuan semua pihak, oleh karena itu penulis mengucapkan banyak terima kasih kepada :

1. Prof. Dr. Fathur Rokhman, M.Hum., Rektor Universitas Negeri Semarang.
2. Prof. Dr. Zaenuri, S.E, M.Si., Akt., Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Semarang.
3. Drs. Arief Agoestanto, M.Si, Ketua Jurusan Matematika FMIPA Universitas Negeri Semarang.
4. Drs. Mashuri, M.Si., Ketua Prodi Matematika Jurusan Matematika FMIPA Universitas Negeri Semarang.
5. Dr. Nur Karomah Dwidayati, M.Si., Dosen Pembimbing I yang telah memberikan bimbingan, pengarahan, nasihat, sata dan dorongan selama penyusunan skripsi ini.
6. Dra. Sunarmi, Dosen Pembimbing II yang telah memberikan bimbingan, pengarahan, nasihat, dan dorongan selama penyusunan skripsi ini.

7. Dr. Scolastika Mariani, M.Si., Dosen Penguji dan Dosen Wali yang telah memberikan penilaian dan saran dalam perbaikan skripsi ini.
8. Dosen jurusan Matematika Universitas Negeri Semarang yang telah membekali dengan berbagai ilmu selama mengikuti perkuliahan sampai akhir penulisan skripsi ini.
9. Bapak dan Ibu tercinta, Bapak Sugiyanto, Alm. Ibu Salminah, dan Ibu Tri Asih, serta keluarga tercinta yang senantiasa memberi ridho, doa, dan dukungan yang tiada putusnya.
10. Teman-teman Matematika UNNES 2014 yang berjuang bersama.
11. Bapak dan Ibu Karyawan Tata Usaha Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Semarang yang telah membantu kebutuhan administrasi.
12. Semua pihak yang telah memberi bantuan dalam penelitian dan penulisan skripsi ini.

Penulis menyadari bahwa dalam penyusunan skripsi ini masih terdapat banyak kekurangan. Oleh karena itu, penulis mengharapkan saran dan kritik yang membangun dari pembaca.

Semarang, September 2018

Penulis

ABSTRAK

Putri, Tiffani Dita Permata. 2018. *Pemodelan Regresi Nonparametrik Spline Multivariabel*. Skripsi. Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Semarang. Pembimbing: Dr. Nur Karomah Dwidayati, M.Si, M.Pd., dan Dra. Sunarmi, M.Pd.

Kata Kunci : Regresi Nonparametrik, *Spline*, Knot, GCV

Analisis regresi merupakan salah satu teknik analisis dalam statistika yang digunakan untuk mengestimasi pola hubungan pasangan data antara variabel respon dan variabel prediktor. Pendekatan regresi yang banyak dikembangkan adalah regresi parametrik dan regresi nonparametrik. Banyak asumsi yang harus terpenuhi pada regresi parametrik. Regresi nonparametrik digunakan apabila asumsi-asumsi pada regresi parametrik tidak terpenuhi

Regresi nonparametrik adalah model regresi yang digunakan untuk mengetahui hubungan antara variabel respon dengan variabel prediktor dengan asumsi bentuk kurva regresi tidak diketahui. Salah satu model yang digunakan untuk mengestimasi kurva regresi nonparametrik adalah *spline*. *Spline* merupakan polinomial sepotong-sepotong yang memiliki sifat tersegmen dan kontinu sehingga memberikan fleksibilitas lebih dibanding dengan polinomial biasa. Dalam penelitian ini, pemilihan titik knot merupakan persoalan penting dalam mengestimasi model karena dapat mempengaruhi nilai error. Titik knot merupakan titik yang menunjukkan perubahan pola data pada sub-sub interval tertentu. Pada penelitian ini, pemilihan titik knot akan menggunakan metode *generalized cross validation* (GCV) dengan memilih nilai GCV terkecil. Data yang digunakan dalam penelitian ini adalah data tingkat kriminalitas tahun 2016 beserta faktor-faktor yang diduga mempengaruhi tingkat kriminalitas. Tujuan dari penelitian ini adalah untuk memperoleh model regresi *spline* linier multivariabel terbaik pada data tingkat kriminalitas.

Hasil penelitian menunjukkan bahwa harapan lama sekolah (X_2) dan persentase penduduk berumur 15 tahun keatas yang tamat SMP (X_5) merupakan faktor yang paling berpengaruh terhadap tingkat kriminalitas di Jawa Tengah (Y) dengan nilai GCV minimum sebesar 0,01231297 dan koefisien determinasi sebesar 97,97%, serta terdapat delapan titik knot yang digunakan dalam model yaitu $K_1 = 13,49$; $K_2 = 13,59$; $K_3 = 13,69$; $K_4 = 13,89$; $K_5 = 8,34$; $K_6 = 8,84$; $K_7 = 16,84$; $K_8 = 17,84$.

DAFTAR ISI

	Halaman
HALAMAN JUDUL	i
HALAMAN PERNYATAAN	iii
HALAMAN PENGESAHAN	iv
MOTTO DAN PERSEMBAHAN	v
KATA PENGANTAR	vii
ABSTRAK	viii
DAFTAR ISI	ix
DAFTAR TABEL	xiii
DAFTAR GRAFIK	xiv
DAFTAR LAMPIRAN	xv
BAB I PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Rumusan Masalah	4
1.3 Batsan Masalah	4
1.4 Tujuan Penelitian	4
1.5 Manfaat Penelitian.....	5
1.6 Sistematika Penulisan	6
1.6.1 Bagian Awal	6
1.6.2 Bagian Isi.....	6

1.6.3 Bagian Akhir	7
BAB II TINJAUAN PUSTAKA	8
2.1 Analisis Regresi	8
2.2 Regresi Parametrik	9
2.3 Regresi Nonparametrik	9
2.4 Regresi <i>Spline</i>	11
2.5 Regresi <i>Spline</i> Multivariabel	13
2.6 Estimasi Parameter Regresi Nonparamaterik <i>Spline</i>	15
2.6.1 <i>Maximim Likelihood Estimation</i> (MLE)	16
2.6.2 <i>Ordinary Least Square</i> (OLS)	17
2.7 Kriteria Pemilihan Model Terbaik	19
2.7.1 <i>Mean Square Error</i>	19
2.7.2 <i>Generalized Cross Validation</i>	20
2.8 Koefisien Determinasi	21
2.9 Pengujian Parameter	22
2.9.1 Uji Serentak	22
2.9.2 Uji Individu	23
2.10 Pengujian Asumsi Residual	24
2.11 <i>Software R</i>	25
2.12 Perhitungan Manual	27
2.13 Tingkat Kriminalitas	33
2.14 Penelitian Yang Relevan	38

2.15 Kerangka Berpikir	41
BAB III METODE PENELITIAN	44
3.1 Identifikasi Masalah	44
3.2 Fokus Penelitian	44
3.3 Jenis dan Sumber Data	45
3.4 Variabel Penelitian	45
3.5 Analisis Data	46
BAB IV HASIL PENELITIAN DAN PEMBAHASAN.....	48
4.1 Hasil Penelitian	48
4.1.1 Ukuran Statistik	48
4.1.2 <i>Scatterplot</i>	52
4.1.3 Pemilihan Titik Knot Optimal.....	53
4.1.3.1 Pemilihan Titik Knot Secara Parsial	54
4.1.3.2 Pemilihan Titik Knot Secara Serentak	69
4.1.4 Pengujian Model Regresi <i>Spline</i> Linier Multivariabel.....	74
4.1.4.1 Pengujian Parameter	75
4.1.4.2 Pengujian Asumsi Residual.....	77
4.1.5 Koefisien Determinasi	81
4.2 Pembahasan.....	82
BAB V PENUTUP	93
5.1 Simpulan.....	93
5.2 Saran.....	94

DAFTAR PUSTAKA	95
LAMPIRAN.....	99

DAFTAR TABEL

Tabel	Halaman
4.1 Ukuran Statistik	48
4.2 Lokasi Knot dan Nilai GCV Variabel X_1	55
4.3 Koefisien Parameter Model regresi Spline Linier Tiga Knot (X_1)	55
4.4 Lokasi Knot dan Nilai GCV Variabel X_2	58
4.5 Koefisien Parameter Model regresi Spline Linier Empat Knot (X_2)	58
4.6 Lokasi Knot dan Nilai GCV Variabel X_3	61
4.7 Koefisien Parameter Model regresi Spline Linier Empat Knot (X_3)	61
4.8 Lokasi Knot dan Nilai GCV Variabel X_4	64
4.9 Koefisien Parameter Model regresi Spline Linier Tiga Knot (X_4)	64
4.10 Lokasi Knot dan Nilai GCV Variabel X_5	67
4.11 Koefisien Parameter Model regresi Spline Linier Empat Knot (X_5)	67
4.12 Titik Knot Optimum dan Nilai GCV Minimum Variabel Prediktor	69
4.13 Kombinasi Tiga Titik Knot Variabel Prediktor	70
4.14 Koefisien Parameter Model Regresi <i>Spline</i> Linier Multivariabel Delapan Titik Knot	71
4.15 Nilai Estimasi Model dan Nilai Residual	72
4.16 Nilai t Hitung Parameter Regresi	76

DAFTAR GAMBAR

Gambar	Halaman
2.1 Kerangka Berpikir.....	43
3.1 Diagram Alir Penelitian	47
4.1 <i>Scatterplot</i> Seluruh Variabel Prediktor terhadap Variabel Respon.....	53
4.2 <i>Scatterplot</i> Tingkat Kriminalitas Di Tiap Kantor Kepolisian.....	54
4.3 Kurva Model Regresi <i>Spline</i> Linier Tiga Knot (X_1)	56
4.4 <i>Scatterplot</i> Tingkat Kriminalitas Pada Harapan Lama Sekolah	57
4.5 Kurva Model Regresi <i>Spline</i> Linier Empat Knot (X_2)	59
4.6 <i>Scatterplot</i> Tingkat Kriminalitas Pada Persentase Penduduk Berumur 15 Tahun Ke Atas yang Belum Pernah Sekolah	60
4.7 Kurva Model Regresi <i>Spline</i> Linier Empat Knot (X_3)	62
4.8 <i>Scatterplot</i> Tingkat Kriminalitas Pada Persentase Penduduk Miskin	63
4.9 Kurva Model Regresi <i>Spline</i> Linier Tiga Knot (X_4)	65
4.10 <i>Scatterplot</i> Tingkat Kriminalitas Pada Persentase Penduduk Berumur 15 Tahun Ke Atas Yang Tamat SMP	66
4.11 Kurva Model Regresi <i>Spline</i> Linier Empat Knot (X_5)	68
4.12 Kurva Y dan \hat{Y}	74
4.13 Plot ACF Residual	79
4.14 Kurva Kenormalan Residual Regresi <i>Spline</i> Linier Multivariabel	80

DAFTAR LAMPIRAN

Lampiran	Halaman
1. Contoh Data Penelitian	100
2. Output Perhitungan Ukuran Statistik	101
3. Data Tingkat Kriminalitas dan Faktor yang Mempengaruhi Menurut Kab/Kota di Jawa tengah Tahun 2016	102
4. Output GCV 1 Titik Knot	104
5. Output GCV 2 Titik Knot	107
6. Output GCV 3 Titik Knot	110
7. Output GCV 4 Titik Knot	113
8. Output GCV 5 Titik Knot	117
9. Output GCV 8 Titik Knot dan 2 Variabel	121
10. Output GCV 12 Titik Knot dan 3 Variabel.....	123
11. <i>Syntax</i> Program <i>R</i>	125
12. Output Uji F	135
13. Output Uji Glejser	136

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Statistika merupakan salah satu ilmu yang memuat sangat banyak cabang ilmu, diantaranya *time series* (runtun waktu), proses stokastik, probabilitas, rancangan percobaan, analisis regresi, reliabilitas, dan lain sebagainya (Budiantara, 2009). Salah satu cabang ilmu yang sangat penting dan populer dikalangan masyarakat adalah analisis regresi. Analisis regresi merupakan salah satu teknik analisis dalam statistika yang digunakan untuk mengestimasi pola hubungan pasangan data antara variabel prediktor dengan variabel respon. Jika nilai variabel prediktor diketahui, maka nilai variabel respon dapat diestimasi dengan menggunakan fungsi regresi (Hardle, 1990:3). Tujuan umum digunakannya analisis regresi adalah untuk meramalkan atau memprediksi nilai dari suatu variabel dalam hubungannya dengan variabel lain berdasarkan persamaan regresi yang diperoleh.

Dewasa ini terdapat tiga jenis model pendekatan regresi yang banyak dikembangkan oleh para peneliti, yaitu pendekatan regresi parametrik, regresi nonparametrik, dan regresi semiparametrik (Budiantara, 2006). Dalam regresi parametrik, pola hubungan antara dua variabel atau lebih umumnya berpola parametrik

seperti linier, kuadrat, kubik dan lainnya. Untuk memodelkan data menggunakan regresi parametrik linier, kuadrat, kubik atau lainnya dimulai dengan menggunakan *scatter plot* (Budiantara, 2006). Dengan memulai *scatter plot* menggunakan beberapa software statistik seperti SPSS, Minitab, Matlab, R, atau yang lainnya dapat diketahui pola datanya. Jika *scatter plot* cenderung mengikuti pola linier maka digunakan model regresi parametrik linier, jika *scatter plot* data cenderung mengikuti pola kuadratik maka digunakan model regresi parametrik kuadratik, dan seterusnya. Regresi nonparametrik digunakan apabila bentuk kurva regresi tidak berpola atau tidak diketahui. Dalam regresi nonparametrik, data akan mencari bentuk estimasi dari kurva regresinya (Eubank, 1988). Artinya pendekatan regresi nonparametrik sangat fleksibel dan sangat objektif. Selain regresi parametrik dan regresi nonparametrik, terdapat regresi semiparametrik. Regresi semiparametrik terbentuk dari komponen model yang diestimasi secara parametrik dan ada komponen lain menggunakan pendekatan nonparametrik. Bisa dikatakan bahwa regresi semiparametrik adalah gabungan antara regresi parametrik dan regresi nonparametrik.

Pendekatan yang sering diteliti saat ini adalah regresi nonparametrik. Terdapat beberapa model pendekatan dalam mengestimasi kurva regresi nonparametrik, seperti *Kernel*, Histogram, Deret Orthogonal, Deret Fourier, Wavelets, dan *Spline* (Budiantara, 2004). Penelitian terdahulu mengenai regresi dengan pendekatan nonparametrik telah diteliti oleh Aydin (2007) tentang perbandingan teknik yang digunakan untuk memprediksi regresi nonparametrik model dengan melakukan *smoothing* estimator regresi *spline* dan regresi *kernel*. Menurut nilai MSE, RMSE, MAE, dan MAPE dan

dilihat dari segi grafis pada data produk nasional bruto, *spline* menunjukkan hasil yang lebih baik dari *kernel*. Penelitian lain dilakukan oleh Ratnasari (2015) membandingkan *spline* dan deret fourier untuk regresi nonparametrik multivariabel pada data kemiskinan di Papua. Dari nilai R^2 dan nilai minimum GCV menunjukkan bahwa metode *spline* lebih baik dari pada metode deret fourier.

Pendekatan *spline* merupakan satu-satunya pendekatan dengan tata cara optimasi yang sangat baik. *Spline* dapat memodelkan data dengan pola berubah-ubah pada sub-sub interval tertentu, dikarenakan *spline* merupakan salah satu jenis potongan polinomial dengan sifat tersegmen. Sifat tersegmen memungkinkan adanya penyesuaian secara lebih efektif terhadap karakteristik lokal suatu data (Budiantara, 2007). Menurut Wahba (1990), *spline* memiliki sifat-sifat statistik yang berguna untuk menganalisis hubungan dalam regresi. Selanjutnya, regresi dengan satu variabel dependen dan lebih dari satu variabel independen disebut sebagai regresi *spline* multivariabel (Budiantara, 2004).

Berdasarkan uraian tersebut penulis ingin mengambil judul “**Pemodelan Regresi Nonparametrik *Spline* Multivariabel**”. Pemodelan akan diaplikasikan pada data tingkat kriminalitas di Jawa Tengah pada tahun 2016 dan dengan menggunakan bantuan *software R* untuk menentukan GCV minimum. Diharapkan dengan adanya penelitian ini dapat mengetahui faktor-faktor yang mempengaruhi tingkat kriminalitas di Jawa Tengah secara signifikan.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, maka permasalahan penelitian ini dirumuskan sebagai berikut.

1. Bagaimana pemodelan regresi *spline* multivariabel pada tingkat kriminalitas di Jawa Tengah?
2. Faktor-faktor apa saja yang mempengaruhi tingkat kriminalitas di Jawa Tengah?
3. Faktor apa yang dominan mempengaruhi tingkat kriminalitas di Jawa Tengah?

1.3 Batasan Masalah

Berdasarkan masalah yang telah dijelaskan, batasan masalah pada penelitian ini hanya mencakup faktor-faktor yang mempengaruhi tingkat kriminalitas di Jawa Tengah pada tahun 2016 seperti jumlah kantor polisi menurut kab/kota di Jawa Tengah, Harapan lama sekolah menurut kab/kota di Jawa Tengah, persentase penduduk berumur 15 tahun keatas yang belum pernah sekolah menurut kab/kota di Jawa Tengah, Persentase penduduk miskin menurut kab/kota di Jawa Tengah, dan Persentase penduduk berumur 15 tahun keatas yang tamat SMP menurut kab/kota di Jawa Tengah.

1.4 Tujuan Penelitian

Berdasarkan permasalahan yang telah dijelaskan, maka tujuan dari penelitian ini adalah sebagai berikut.

1. Mengetahui model regresi *spline* multivariabel pada tingkat kriminalitas di Jawa Tengah.
2. Mengetahui faktor-faktor yang mempengaruhi tingkat kriminalitas di Jawa Tengah.
3. Mengetahui faktor yang dominan mempengaruhi angka kriminalitas di Jawa Tengah.

1.5 Manfaat Penelitian

Manfaat yang akan diperoleh dari penelitian ini adalah sebagai berikut.

1. Bagi penulis

Manfaat yang diperoleh bagi penulis adalah penambahan wawasan tentang pemodelan regresi nonparametrik *spline* multivariabel pada data tingkat kriminalitas yaitu mendapatkan model terbaik untuk memprediksi faktor yang mempengaruhi tingkat kriminalitas serta mengetahui faktor yang dominan mempengaruhi tingkat kriminalitas di Jawa Tengah.

2. Bagi pembaca

Manfaat yang diperoleh dari penelitian ini bagi pembaca khususnya mahasiswa adalah sebagai referensi guna melakukan penelitian lain sebagai pengembangan model regresi nonparametrik *spline* multivariabel lainnya.

1.6 Sistematika Penulisan

Sistematika penulisan dalam skripsi ini terdiri dari 3 (tiga) bagian, yaitu bagian awal, bagian isi, dan bagian akhir.

1.6.1 Bagian Awal

Bagian awal terdiri dari halaman judul, halaman pengesahan, halaman motto dan persembahan, halaman kata pengantar, halaman abstrak, halaman daftar isi, halaman daftar tabel, halaman daftar gambar, dan halaman daftar lampiran.

1.6.2 Bagian Isi

Bagian isi terdiri dari 5 (lima) bab. Adapun bab tersebut diuraikan seperti berikut.

Bab I Pendahuluan

Pada bab pendahuluan berisi latar belakang masalah, rumusan masalah, pembatasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

Bab II Tinjauan Pustaka

Pada bab tinjauan pustaka berisi landasan teori seperti regresi, regresi nonparametrik, pengujian asumsi, dll

Bab III Metode Penelitian

Pada bab metode penelitian berisi identifikasi masalah, focus penelitian, jenis dan sumber data, analisis dan pemecahan masalah, dan penarikan kesimpulan.

Bab IV Hasil dan Pembahasan

Pada bab hasil dan pembahasan berisi hasil penelitian beserta pembahasannya sebagai jawaban atas permasalahan.

Bab V Penutup

Pada bab penutup berisi tentang simpulan dari pembahasan penelitian dan saran dari kesimpulan

1.6.3 Bagian Akhir

Bagian akhir meliputi daftar pustaka dan lampiran-lampiran yang mendukung penelitian.

BAB 2

TINJAUAN PUSTAKA

2.1 Analisis Regresi

Regresi merupakan model analisis statistik yang umumnya digunakan untuk mencari hubungan sebab akibat antara variabel yang satu dengan variabel yang lain. Analisis regresi merupakan analisis statistika yang digunakan untuk menyelidiki pola hubungan fungsional antara variabel respon dan variable prediktor (Budiantara, 2009). Hubungan fungsional antarvariabel tersebut dijelaskan dalam sebuah kurva yang disebut dengan kurva regresi.

Analisis regresi menurut jenis variable respon dan prediktor dibagi menjadi 2 (dua) macam yaitu regresi univariabel dan multivariabel. Regresi univariabel dilakukan untuk mencari hubungan satu variabel respon dan satu variabel prediktor. Sedangkan regresi multivariabel dilakukan apabila jumlah variabel respon lebih dari satu dan variabel prediktor lebih dari satu.

Pemodelan menggunakan analisis regresi akan menghasilkan suatu persamaan yang disebut dengan persamaan regresi. Misal diberikan data berpasangan $(t_i, y_i), i = 1, \dots, n$ dan pola hubungan antara t_i dengan y_i diasumsikan mengikuti model regresi sebagai berikut (Eubank, 1999, p. 1).

$$y_i = f(t_i) + \varepsilon_i, i = 1, 2, \dots, n \quad (2.1)$$

Dengan y_i adalah variabel respon ke- i , $f(t_i)$ adalah fungsi f yang tidak diketahui nilainya pada pengamatan t_1, t_2, \dots, t_n , dan ε_i merupakan galat yang diasumsikan saling bebas dengan mean nol dan variansi σ^2 . Fungsi f menyatakan fungsi regresi atau kurva regresi (Hardle, 1994).

Pada umumnya, terdapat dua pendekatan untuk mengestimasi kurva regresi, yaitu pendekatan parametrik dan pendekatan nonparametrik. Pendekatan parametrik dilakukan jika pola data diketahui seperti berpola linier, kuadrat, kubik, dan lainnya. Jika tidak diketahui pola data, maka digunakan pendekatan nonparametrik.

2.2 Regresi Parametrik

Bentuk fungsi f dari (2.1) dalam regresi parametrik diketahui dan asumsi kenormalan distribusi galat harus terpenuhi. Model persamaan regresi parametrik linear dengan p variabel pada umumnya dituliskan sebagai berikut.

$$y_i = \beta_0 + \sum_{j=1}^p \beta_j x_{ij} + \varepsilon_i \quad (2.2)$$

dengan x_{ij} menyatakan nilai variabel prediktor ke- j untuk pengamatan ke- i dengan $j = 1, 2, \dots, p$, $i = 1, 2, \dots, n$ dan $\beta_0, \beta_1, \dots, \beta_p$ adalah parameter regresi.

2.3 Regresi Nonparametrik

Regresi nonparametrik merupakan model regresi yang digunakan untuk mengetahui hubungan antara variabel respon dengan variabel prediktor dengan asumsi

bentuk kurva regresi tidak diketahui. Ide dasar nonparametric adalah menggunakan data untuk menyimpulkan suatu kuantitas yang tidak diketahui dengan sedikit asumsi (Wasserman, 2006). Dalam kasus nonparametrik pemenuhan asumsi kenormalan galat tidak diperlukan, karena penggunaan regresi nonparametrik tidak terikat akan asumsi-asumsi seperti yang mendasari statistika parametrik. Statistika nonparametrik merupakan statistika bebas distribusi. Regresi nonparametrik hanya memiliki asumsi bahwa kurva regresi yang dibentuk bersifat *smooth* atau mulus. Regresi nonparametrik mempunyai keunggulan dibandingkan regresi parametrik, yaitu mempunyai tingkat fleksibilitas yang lebih tinggi, artinya data mampu mencari bentuk estimasi kurva regresi secara mandiri tanpa dipengaruhi faktor dari peneliti (Eubank, 1988: 3). Terdapat beberapa model regresi nonparametrik yang sering digunakan seperti *spline*, *MARS*, *kernel*, *wavelet*, dan *deret fourier*. Bentuk model regresi nonparametrik secara umum dapat ditulis sebagai berikut (Eubank, 1999).

$$y_i = f(t_i) + \varepsilon_i ; i = 1, 2, \dots, n \quad (2.3)$$

dimana:

y_i : variabel respon pengamatan ke- i .

t_i : variabel prediktor pengamatan ke- i .

$f(t_i)$: fungsi regresi pada pengamatan ke- i .

ε_i : residual atau *error* untuk pengamatan ke- i dengan $i = 1, 2, \dots, n$. ε diasumsikan berdistribusi $N(0, \sigma^2)$. ε_i bisa ditulis $(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)^2$

2.4 Regresi *Spline*

Salah satu metode regresi nonparametrik adalah *spline*. *Spline* merupakan polinomial sepotong-sepotong. Fungsi *spline* memiliki sifat tersegmen yang kontinu. Sifat ini memberikan fleksibilitas lebih dibandingkan dengan polinomial biasa (Eubank, 1988: 196). *Spline* juga merupakan generalisasi dari polinomial. Fungsi *spline* berorde q dengan titik-titik knot K_1, K_2, \dots, K_M didefinisikan sebagai sembarang fungsi f yang disajikan dalam bentuk (Eubank, 1988: 89):

$$f(x) = \beta_0 + \sum_{j=1}^q \beta_j x^j + \sum_{k=1}^M \beta_{j+k} (x - K_k)_+^q \quad (2.4)$$

dengan fungsi potongan (*truncated*) dengan satu titik knot sebagai berikut (Tripena, 2011):

$$f_1(x) = \begin{cases} \beta_0 + \beta_1 x & , x < K \\ \beta_0 + \beta_1 x + \beta_2 (x - K) & , x \geq K \end{cases} \quad (2.5)$$

Fungsi *spline* dengan empat titik knot pada ($x = K_1, x = K_2, x = K_3, x = K_4$) yaitu

$$f_4(x) = \beta_0 + \beta_1 x + \beta_2 (x - K_1)_+^1 + \beta_3 (x - K_2)_+^1 + \beta_4 (x - K_3)_+^1 + \beta_5 (x - K_4)_+^1$$

Fungsi $f_4(x)$ dapat disajikan dalam bentuk:

$$f_4(x) = \begin{cases} \beta_0 + \beta_1 x & , x < K_1 < K_2 < K_3 < K_4 \\ \beta_0 + \beta_1 x + \beta_2 (x - K_1) & , K_1 \leq x < K_2 < K_3 < K_4 \\ \beta_0 + \beta_1 x + \beta_2 (x - K_1) + \beta_3 (x - K_2) & , K_2 \leq x < K_3 < K_4 \\ \beta_0 + \beta_1 x + \beta_2 (x - K_1) + \beta_3 (x - K_2) + \beta_4 (x - K_3) & , K_3 \leq x < K_4 \\ \beta_0 + \beta_1 x + \beta_2 (x - K_1) + \beta_3 (x - K_2) + \beta_4 (x - K_3) + \beta_5 (x - K_4) & , x \geq K_4 \end{cases}$$

(2.6)

dengan

β_j : parameter model

β_0 : intersep

β_{j+k} : slope pada peubah x truncated knot ke- k pada *spline* berorde q

(konstanta riil).

x_i : variabel prediktor.

K_k : titik knot ke- k .

M : banyaknya knot dalam variabel respon ke- j .

Spline adalah jumlahan dari fungsi polinomial berderajat q dengan *truncated* derajat q . Model regresi nonparametrik *spline* orde ke- m dapat disajikan dalam bentuk sebagai berikut.

$$y_i = \beta_0 + \sum_{j=1}^q \beta_j x_i^j + \sum_{k=1}^M \beta_{j+k} (x_i - K_k)^q + \varepsilon_i; \quad i = 1, 2, \dots, n \quad (2.7)$$

Dengan keterangan sebagai berikut:

y_i : variabel respon pada pengamatan ke- i

β_0 : intersep

β_j : parameter polinomial pada pengamatan ke- j

x_i : variabel predictor pada pengamatan ke- i

x_i^j : variabel predictor pada pengamatan ke- i berorde j

β_{j+k} : parameter terpotong bernilai riil pada pengamatan ke- $(j+k)$

K_k : titik knot ke- k

ε_i : galat pada pengamatan ke- i

M : banyaknya titik knot

Berdasarkan bentuk matematis fungsi *spline* tersebut, dapat dikatakan bahwa *spline* merupakan model polinomial yang sepotong-sepotong (*piecewise polynomial*). Hanya saja, *spline* masih bersifat kontinu pada knot-knotnya berupa potongan polinomial mulus yang masih memungkinkan memiliki sifat tersegmen (Eubank, 1988: 196). Sehingga kurva yang dibentuk tersegmen pada titik tersebut dan pada setiap fungsi $f(x)$. Knot diartikan sebagai suatu titik fokus dalam fungsi *spline*.

Untuk mengestimasi kurva regresi nonparametrik dengan pendekatan *spline polynomial truncated*, maka secara teoritis dapat dilakukan dengan mencari model *spline polynomial truncated* terbaik berdasarkan titik knot yang optimum yaitu banyak titik knot dan letak titik-titik knot tersebut.

2.5 Regresi *Spline* Multivariabel

Model regresi *spline* multivariabel yaitu model regresi *spline* yang mempunyai lebih dari satu variabel prediktor. Karena variabel prediktor lebih dari satu, maka dapat ditulis kembali dalam bentuk:

$$y_i = \sum_{i=1}^p f(x_{ii}) + \varepsilon_i \quad (2.8)$$

Persamaan (2.7) dihipotesis dengan fungsi *spline*, sehingga dapat ditulis kembali dalam bentuk:

$$y_i = \beta_0 + \sum_{j=1}^q \beta_{ji} x_{li}^j + \sum_{k=1}^M \beta_{j+kl} (x_{li} - K_{kl})_+^q + \varepsilon_i \quad (2.9)$$

Persamaan tersebut dapat ditulis dalam bentuk matriks sebagai berikut:

$$\begin{aligned}
\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} &= \begin{bmatrix} \mathbf{1} & x_{11} & \cdots & x_{11}^q & (x_{11} - K_{11})_+^q & \cdots & (x_{11} - K_{M1})_+^q \\ \mathbf{1} & x_{12} & \cdots & x_{12}^q & (x_{12} - K_{11})_+^q & \cdots & (x_{12} - K_{M1})_+^q \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{1} & x_{1n} & \cdots & x_{1n}^q & (x_{1n} - K_{11})_+^q & \cdots & (x_{1n} - K_{M1})_+^q \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_{11} \\ \vdots \\ \beta_{q1} \\ \beta_{j+1.1} \\ \vdots \\ \beta_{j+M.1} \end{bmatrix} + \cdots \\
\begin{bmatrix} \mathbf{1} & x_{p1} & \cdots & x_{p1}^q & (x_{p1} - K_{1p})_+^q & \cdots & (x_{p1} - K_{Mp})_+^q \\ \mathbf{1} & x_{p2} & \cdots & x_{p2}^q & (x_{p2} - K_{1p})_+^q & \cdots & (x_{p2} - K_{Mp})_+^q \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{1} & x_{pn} & \cdots & x_{pn}^q & (x_{pn} - K_{1p})_+^q & \cdots & (x_{pn} - K_{Mp})_+^q \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_{1p} \\ \vdots \\ \beta_{qp} \\ \beta_{j+1.p} \\ \vdots \\ p \end{bmatrix} + \varepsilon_i \quad (2.10)
\end{aligned}$$

Persamaan dapat disederhanakan menjadi:

$$y = W(K_{11}, \dots, K_{M1}, \dots, K_{1p}, \dots, K_{Mp})\theta + \varepsilon \quad (2.11)$$

Dengan keterangan sebagai berikut:

$Y = [Y_1, Y_2, \dots, Y_n]^T$: vektor variabel respon berukuran $n \times 1$.

$\varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)^T$: vektor galat berukuran $n \times 1$.

$\theta = (\beta_0, \beta_{11}, \dots, \beta_{q1}, \beta_{j+1.1}, \dots, \beta_{j+M.1}, \dots, \beta_0, \beta_{1p}, \dots, \beta_{qp}, \beta_{j+1.p}, \dots, \beta_{j+M.p})^T$

adalah vektor parameter berukuran $(pq + M_1 + M_2 + \dots + M_p) \times 1$

$$W(K_{11}, \dots, K_{M1}, \dots, K_{1p}, \dots, K_{Mp}) = \begin{bmatrix} \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{1} \\ x_{11} & x_{12} & \dots & x_{1n} \\ \vdots & \vdots & \dots & \vdots \\ x_{11}^q & x_{12}^q & \dots & x_{1n}^q \\ (x_{11} - K_{11})_+^q & (x_{12} - K_{11})_+^q & \dots & (x_{1n} - K_{11})_+^q \\ \vdots & \vdots & \dots & \vdots \\ (x_{11} - K_{M1})_+^q & (x_{12} - K_{M1})_+^q & \dots & (x_{1n} - K_{M1})_+^q \\ \vdots & \vdots & \dots & \vdots \\ x_{p1} & x_{p2} & \dots & x_{pn} \\ \vdots & \vdots & \dots & \vdots \\ x_{p1}^q & x_{p2}^q & \dots & x_{pn}^q \\ (x_{p1} - K_{1p})_+^q & (x_{p2} - K_{1p})_+^q & \dots & (x_{pn} - K_{1p})_+^q \\ \vdots & \vdots & \dots & \vdots \\ (x_{p1} - K_{Mp})_+^q & (x_{p2} - K_{Mp})_+^q & \dots & (x_{pn} - K_{Mp})_+^q \end{bmatrix}^T$$

adalah matriks berukuran $n \times (pq + M_1 + M_2 + \dots + M_p)$.

2.6 Estimasi Parameter Regresi Nonparametrik *Spline*

Metode yang umumnya digunakan untuk mengestimasi parameter regresi nonparametrik pada *spline* adalah *maximum likelihood estimation* (MLE) dan *ordinary least square* (OLS). Performansi keduanya dalam menghasilkan estimasi parameter sama baiknya, namun yang sering digunakan adalah metode *ordinary least square*.

2.6.1 Maximum Likelihood Estimation (MLE)

Estimasi parameter regresi nonparametrik *spline* dapat diperoleh dengan menggunakan metode *Maximum Likelihood Estimation* (MLE) berdasarkan asumsi yaitu t_i dan ε_i pada persamaan (2.4) merupakan variabel acak yang saling bebas, dan $\varepsilon_i \sim N(0, \sigma^2)$. Oleh karena itu, y_i juga berdistribusi normal dengan *mean* $f(t_i)$ dan variansi σ^2 (Mubarak, 2012). Sehingga fungsi densitas peluang y_i menjadi

$$f(y; f(t), \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(y-f(t))^2}{2\sigma^2}\right], f(t) > 0, \sigma^2 > 0 \quad (2.12)$$

Selanjutnya fungsi likelihood dapat dituliskan sebagai berikut.

$$L(y, f) = \prod_{i=1}^n f(y; f(t), \sigma^2) \quad (2.13)$$

$$L(y, f) = \prod_{i=1}^n \frac{\exp\left(-\frac{1}{2\sigma^2}(y_i - f(t_i))^2\right)}{\sqrt{2\pi\sigma^2}}$$

$$L(y, f) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - f(t_i))^2\right) \quad (2.14)$$

Dengan memaksimalkan fungsi likelihood $L(y, f)$ akan diperoleh estimasi titik untuk fungsi f sebagai berikut.

$$\begin{aligned} \max_{f, j=0} \{L(y, f)\} &= \max_{\beta \in R^{m+n}} \{(2\pi\sigma^2)^{-n/2} \exp \\ &\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \sum_{j=0}^m \beta_j t_i^j - \sum_{l=1}^n \beta_{m+l} (t_i - k_l)^m)\right)^2 \} \end{aligned} \quad (2.15)$$

Kemudian dilakukan transformasi logaritma sehingga menghasilkan persamaan sebagai berikut.

$$\begin{aligned} \log L(y, \beta) &= -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n \\ &(y_i - \sum_{j=0}^m \beta_j t_i^j - \sum_{l=1}^n \beta_{m+l} (t_i - k_l)^m)^2 \end{aligned} \quad (2.16)$$

Misalkan $x_i(k) = [1 \ t_i \ t_i^2 \ \dots \ t_i^m (t_i - k_1)^m + (t_i - k_2)^m + \dots (t_i - k_n)^m]$

Dan

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{m+n} \end{bmatrix}$$

Maka

$$\log L(y, \beta) = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - x_i(k)\beta)^2 \quad (2.17)$$

Dengan $k = (k_1, k_2, \dots, k_n)^T$

Selanjutnya diturunkan secara parsial terhadap β dan disamakan dengan nol di sisi kanannya, sehingga diperoleh hasil sebagai berikut.

$$\frac{\partial \left(-\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} (Y - X(k)\beta)^T (Y - X(k)\beta) \right)}{\partial \beta} = 0 \quad (2.18)$$

Hasil estimasi parameter

$$\hat{\beta}(k) = (X(k)^T X(k))^{-1} X(k)^T y \quad (2.19)$$

2.6.1 Ordinary Least Square (OLS)

Metode *maximum likelihood estimation (MLE)* bekerja dengan memaksimalkan fungsi *likelihood* $L(\theta)$ atau fungsi *log-likelihood* $\ln L(\theta)$. Selain *maximum likelihood estimation (MLE)*, metode yang biasa digunakan untuk menaksir parameter adalah *ordinary least square (OLS)*. Prinsip dari metode ini adalah dengan meminimumkan galat (*error*) yang dihasilkan oleh model sehingga diharapkan model

regresi menjelaskan data dengan baik. Mengingat galat yang dihasilkan model regresi pada setiap amatan dapat bernilai negatif dan positif, maka untuk menghindari penjumlahan yang bernilai negatif, dicari jumlah dari kuadrat nilai galat.

Berdasarkan persamaan bentuk umum regresi dapat diambil bentuk hubungan baru yaitu sebagai berikut:

$$\boldsymbol{\varepsilon} = Y - W\boldsymbol{\beta} \quad (2.20)$$

sesuai definisi metode kuadrat terkecil, maka:

$$L = \boldsymbol{\varepsilon}\boldsymbol{\varepsilon} = (Y - W\boldsymbol{\beta})(Y - W\boldsymbol{\beta}) \quad (2.21)$$

Asumsikan $YX\boldsymbol{\beta} = X\boldsymbol{\beta}Y$, jadi dapat ditulis menjadi

$$L = YY - \boldsymbol{\beta}WY - \boldsymbol{\beta}WY + \boldsymbol{\beta}\boldsymbol{\beta}WW \quad (2.22)$$

Karena terdapat matriks $\boldsymbol{\beta}XY$ yang bernilai sama, sehingga diperoleh

$$L = YY - 2\boldsymbol{\beta}WY + \boldsymbol{\beta}^2WW \quad (2.23)$$

untuk mendapatkan nilai terkecil dari fungsi L yang bersifat kuadratik, maka gunakan turunan terhadap parameter yang diduga ($\boldsymbol{\beta}$) disamakan dengan nol.

$$\frac{dL}{d\boldsymbol{\beta}} = -2WY + 2\boldsymbol{\beta}WW = 0 \quad (2.24)$$

$$W\boldsymbol{\beta} = Y \quad (2.25)$$

jika ruas kanan dan kiri dikalikan dengan W^T diperoleh

$$W^TW\boldsymbol{\beta} = W^TY \quad (2.26)$$

Asumsikan W^TW mempunyai rank penuh, sehingga W mempunyai invers,

$$(W^TW)^{-1}W^TW\boldsymbol{\beta} = (W^TW)^{-1}W^TY \quad (2.27)$$

karena $(W^TW)^{-1}W^TW = I$ dan $I\boldsymbol{\beta} = \boldsymbol{\beta}$ sehingga diperoleh:

$$B = \hat{\beta} = (W^T W)^{-1} W^T Y \quad (2.28)$$

2.7 Kriteria Pemilihan Model Terbaik

Semakin kecil nilai *error* yang dihasilkan, maka estimasi model akan semakin baik. Pada regresi *spline*, pemilihan jumlah titik knot serta penempatan titik knot yang optimal dapat mempengaruhi nilai *error* dari model yang diperoleh. Menurut Eubank (1988: 32), salah satu metode untuk menentukan titik knot yang optimal adalah *mean square error (MSE)* dan *generalized cross validation (GCV)*.

2.7.1 Mean Square Error

Mean Square Error (MSE) adalah nilai taksiran dari variansi residual. *MSE* juga diartikan sebagai harapan nilai kuadrat perbedaan antara estimator dengan parameter populasi (Vitaningrum, 2018). Model yang baik adalah model dengan *MSE* yang minimum, yang berarti nilai taksiran dari model mendekati nilai sebenarnya. Ukuran kinerja atas estimator yang sederhana adalah kuadrat dari sisaan yang dirata-rata. Rumus *MSE* dapat ditulis sebagai berikut (Eubank, 1988: 24):

$$\begin{aligned} MSE &= n^{-1} \sum_{i=1}^n (Y_i - f(x_i))^2 \\ &= n^{-1} ((I - G)Y)^T ((I - G)Y) \\ MSE &= n^{-1} \sum_{i=1}^n (Y_i - f(x_i))^2 \\ &= n^{-1} (Y - f(x))^T (Y - f(x)) \end{aligned} \quad (2.30)$$

$$= n^{-1}((I - G)Y)^T ((I - G)Y) \quad (2.31)$$

dengan,

n : jumlah data amatan.

Y_i : variabel respon pada pengamatan ke- i .

x_i : variabel prediktor pada pengamatan ke- i .

$f(x_i)$: fungsi regresi pada pengamatan ke- i dengan titik knot.

G : $W(W^T W)^{-1}W^T$

2.7.2 Generalized Cross Validation

Menurut Eubank (1988: 30), *GCV* merupakan modifikasi dari *cross validation* (*CV*). *Cross validation* (*CV*) merupakan suatu metode untuk memilih model berdasarkan pada kemampuan prediksi dari model tersebut. Bentuk model dari *CV* yaitu sebagai berikut:

$$CV = n^{-1} \sum_{i=1}^n \left(\frac{y_i - f(x_i)}{1 - g_{ii}} \right)^2 \quad (2.32)$$

dengan g_{ii} adalah elemen diagonal ke- i dari matriks G .

Persamaan *GCV* diperoleh dengan mengganti g_{ii} pada persamaan dengan $\sum_{i=1}^n g_{ii} = n^{-1}Tr(I - G)$. Nilai dari $Tr(I - G)$ adalah penjumlahan elemen diagonal matriks $(I - G)$. Fungsi *GCV* didefinisikan sebagai :

$$GCV = n^{-1} \frac{\sum_{i=1}^n (y_i - f(x_i))^2}{(n^{-1}Tr(I - G))^2} \quad (2.33)$$

$$GCV = \frac{MSE}{(n^{-1}Tr(I-G))^2} \quad (2.34)$$

dengan $n^{-1}Tr(I - G) < n$ dan $G = W(W^T W)^{-1}W^T$.

Menurut Eubank (1988: 30), GCV dapat digunakan untuk menyelesaikan permasalahan penyelesaian pemilihan parameter untuk meminimalkan MSE atau metode lain pada variansi yang sangat luas. Ketika kriteria GCV memiliki nilai yang minimum, maka model regresi *spline* dapat dikatakan memiliki *error* yang kecil juga.

2.8 Koefisien Determinasi

Salah satu tujuan regresi adalah mendapatkan model terbaik yang mampu menjelaskan hubungan antara prediktor dan variabel respon. Kriteria yang dapat digunakan dalam pemilihan model terbaik salah satunya adalah dengan menggunakan koefisien determinasi atau *R-Square* (R^2). Koefisien determinasi pada [regresi linear](#) sering diartikan sebagai seberapa besar kemampuan semua variabel bebas dalam menjelaskan [varians](#) dari variabel terikatnya. Secara sederhana koefisien determinasi dihitung dengan mengkuadratkan koefisien korelasi (R). Secara umum, semakin besar nilai R^2 , maka semakin baik pula model yang didapatkan. Koefisien determinasi didefinisikan sebagai berikut:

$$R^2 = \frac{SSR}{SST} \quad (2.35)$$

dimana Sum of Square Regression (SSR) dan Sum of Square Total (SST) dirumuskan dengan

$$SSR = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 \quad (2.36)$$

$$SST = \sum_{i=1}^n (y_i - \bar{y})^2 \quad (2.37)$$

Sehingga persamaan dapat dituliskan kembali menjadi:

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (2.38)$$

Besaran nilai R^2 tidak pernah negatif. (Gujarati, 2003)

Menghitung nilai koefisien determinasi merupakan salah satu rangkaian penting dalam menganalisis model data berpasangan melalui regresi. Apabila dengan hanya satu peubah bebas (X) diperoleh nilai koefisien determinasi yang cukup besar maka cukup kita berbicara model RLS saja. Tetapi apabila nilai koefisien determinasinya kecil maka dianjurkan untuk menggunakan metode regresi berganda atau melakukan transformasi data. (Soleh, 2005)

2.9 Pengujian Parameter

Pengujian parameter dimaksudkan untuk menentukan variabel prediktor yang berpengaruh secara simultan terhadap variabel respon. Terdapat dua tahap pengujian yaitu sebagai berikut.

2.9.1 Uji Serentak

Pengujian serentak dilakukan untuk mengetahui apakah parameter model regresi telah signifikan secara bersama-sama. Uji serentak menggunakan hipotesis sebagai berikut.

$$H_0: \beta_1 = \beta_2 = \beta_3 = \dots = \beta_{q+m} = 0$$

H_1 : minimal terdapat satu $\beta_k \neq 0$ dimana $k = 1, 2, \dots, q + m$.

Nilai $(q + m)$ merupakan banyak parameter dalam model regresi nonparametric spline kecuali β_0 .

Statistik uji yang digunakan adalah statistic uji F.

$$F_{hitung} = \frac{MSR}{MSE} \quad (2.39)$$

H_0 ditolak apabila $F_{hitung} > F_{\alpha; (s, n-(q+m)-1)}$ atau $p - value < \alpha$.

2.9.2 Uji Individu

Pengujian parameter model secara individu dilakukan untuk mengetahui apakah parameter secara individu mempunyai pengaruh yang signifikan terhadap variabel respon. Hipotesis yang digunakan adalah sebagai berikut.

$$H_0: \beta_k = 0$$

$$H_1: \beta_k \neq 0, k = 1, 2, \dots, q + m$$

Pengujian secara individu yang digunakan adalah statistik uji t.

$$t_{hitung} = \frac{\widehat{\beta}_k}{\sqrt{var(\beta_k)}} \quad (2.40)$$

Dengan $var(\beta_k) = diag[(X'X)^{-1}\widehat{\sigma}^2]_k$

Dimana $\widehat{\sigma}^2$ merupakan MSE. H_0 ditolak apabila $|t_{hitung}| > t_{\frac{\alpha}{2}; (n-(q+m)-1)}$ atau $p - value < \alpha$.

2.10 Pengujian Asumsi Residual

Pada analisis regresi nonparametrik spline truncated, residual yang terbentuk harus memenuhi asumsi IIDN.

1. Uji Identik

Untuk mengetahui apakah residual identik atau tidak dapat dilakukan dengan uji secara visual dan uji Glejser. Uji secara visual melihat plot antara nilai dugaan respon dengan residual. Apabila tidak terdapat pola pada plot maka tidak terjadi heteroskedastisitas. Untuk mengetahui terjadinya heteroskedastisitas dilakukan dengan menggunakan Uji Glejser. Uji Glejser dilakukan dengan meregresikan nilai mutlak dari residual dengan variabel prediktor yang signifikan terhadap model. Dalam uji glejser dapat dilihat dari tabel ANOVA. Jika nilai $p - value > \alpha$ maka H_0 diterima sehingga tidak terjadi heteroskedastisitas. Nilai $p - value$ menunjukkan taraf kritik terkecil untuk menolak H_0 . Hal ini menunjukkan residual telah memenuhi asumsi identik. Jika asumsi terlanggar atau pada kondisi heteroskedastisitas, maka varians residual tidak konstan sehingga menyebabkan estimasi koefisien kurang akurat (Budiantara, 2014).

2. Uji Independen

Pengujian independen residual bertujuan untuk mengetahui apakah terdapat korelasi antar residual. Korelasi antar residual yaitu korelasi antara residual pada pengamatan ke-I dengan pengamatan ke-(i-1) atau biasanya disebut

dengan autokorelasi (Budiantara, 2014). Untuk mengetahui apakah residual independen atau tidak dapat dilakukan dengan melihat plot ACF dari residual dengan hipotesis sebagai berikut.

$$H_0: \rho = 0 \text{ (residual independen)}$$

$$H_1: \rho \neq 0 \text{ (residual tidak independen)}$$

Apabila terdapat minimal satu autokorelasi pada lag yang keluar batas signifikansi maka H_0 ditolak, yaitu residual tidak independen.

3. Uji Distribusi Normal

Uji normalitas digunakan untuk mengetahui residual telah mengikuti pola distribusi normal. Untuk melakukan pengujian distribusi normal dapat dilakukan dengan Uji Kolmogorov-Smornov dengan hipotesis sebagai berikut,

$$H_0: F_0(x) = F(x) \text{ (residual berdistribusi normal)}$$

$$H_1: F_0 \neq F(x) \text{ (residual tidak berdistribusi normal)}$$

$$\alpha : 0,05$$

Dengan statistik uji $D_{hit} = \sup x |Fn(x) - F_0(x)|$

Tolak H_0 , jika nilai $D_{hit} < D_\alpha$ diterima artinya residual telah berdistribusi normal.

2.11 *Software R*

Software R merupakan salah satu aplikasi statistika yang tersedia secara gratis yang dikembangkan oleh Ross Ihaka dan Robert Gentleman berdasarkan Bahasa S

yaitu sebuah Bahasa komputasi untuk analisis data dan grafik pada tahun 1990an. Pengembangan dan pendistribusian *R* adalah merupakan tugas *dari the R Development Core Team* yang dikelompoki oleh beberapa ahli statistika (Suhartono,2008).

Terdapat kelebihan pada *software R*, yaitu diantaranya serbaguna (*versatile*), interaktif (*interactive*), Berbasis S yaitu turunan dari *tool statistic* komersial S-Plus, Populer.

1. Serbaguna (*versatile*)

R adalah bahasa pemrograman, sehingga tidak ada batasan bagi pengguna untuk memakai prosedur yang hanya terdapat pada paket-paket yang standar. Bahkan pemrograman *R* adalah berorientasi objek dan memiliki banyak *library* yang sangat bermanfaat yang dikembangkan oleh kontributor. Pengguna menambah dan mengurangi *library* tergantung kebutuhan. *R* juga memiliki *interface* pemrograman C, python, bahkan java yang tentu saja berkat jerih payah kontributor aktif proyek *R*. jadi selain bahasa *R* ini cukup pintar, penggunanyapun bisa menjadi lebih pintar dan kreatif. Beberapa analisis yang membutuhkan fungsi lanjutan memang ada yang belum tersedia dalam *R*. tidak berarti *R* tidak menyediakan fasilitas tersebut, namun lebih karena faktor waktu.

2. Interaktif (*interactive*)

Pada saat ini analisis data membutuhkan pengoperasian yang interaktif. Apalagi jika data yang dianalisis adalah data yang bergerak. *R* dilengkapi dengan konektivitas ke *database server*, *olap*, maupun format data *web service* seperti XML, *spreadsheet*

dan sebagainya. Sehingga apabila data set berubah hasil analisa pun dapat segera ikut berubah (*real time*).

3. Berbasis S yaitu turunan dari tool statistic komersial S-Plus

R hampir seluruhnya kompatibel dengan S-Plus. Artinya sebagian besar kode program yang dibuat oleh S dapat dijalankan di S-Plus kecuali fungsi-fungsi yang sifatnya *add-on packages* atau tambahan yang dibuat oleh kontributor proyek R.

4. Populer

Secara umum SAS adalah *software* statistika komersial yang populer, namun demikian R atau S adalah bahasa yang paling populer digunakan oleh peneliti dibidang statistika. Beberapa tulisan berupa jurnal statistika mengkonfirmasi kebenaran hal ini. R juga populer untuk aplikasi kuantitatif dibidang keuangan.

Secara fair Ihaka dan Gentleman menyebutkan kekurangan dari R adalah tidak mudah untuk dipelajari. Beberapa *requirement* awal diperlukan sebelum memperoleh manfaat dari R seperti pemahaman tentang dasar–dasar pemrograman. Namun menurut pendapat penulis, *user* linux seharusnya sudah memiliki pemahaman tentang dasar–dasar pemrograman sehingga akan lebih mudah dan produktif dalam mengeksplorasi tools statistika ini.

2.12 Perhitungan Manual

Perhitungan manual diperlukan agar pembaca mudah untuk memahami langkah-langkah pengerjaan *spline*. Berikut akan disajikan perhitungan manual dalam

mencari nilai *generalized cross validation* (GCV) minimum pada *spline* multivariabel. Data yang akan digunakan adalah data kemiskinan (y), data persentase individu usia sekolah yang bekerja (x_1), dan data persentase rumah tangga dengan pembuangan akhir tinja selain tangka atau saluran (x_2) 10 kecamatan di Kabupaten Banyumas tahun 2011 yang disajikan pada lampiran 1.

- Langkah pertama adalah membuat matriks kolom tiap variabel.

$$y = \begin{bmatrix} 23,82 \\ 15,72 \\ 15,91 \\ 14,92 \\ 18,38 \\ 15,84 \\ 18,28 \\ 16,09 \\ 22,34 \\ 17,67 \end{bmatrix}, x_1 = \begin{bmatrix} 1,18 \\ 0,39 \\ 0,55 \\ 0,33 \\ 0,6 \\ 0,59 \\ 0,72 \\ 1,04 \\ 0,49 \\ 0,95 \end{bmatrix}, x_2 = \begin{bmatrix} 42,44 \\ 25,19 \\ 18,9 \\ 22,99 \\ 27,51 \\ 24,21 \\ 22,62 \\ 23,99 \\ 28,33 \\ 52,74 \end{bmatrix}$$

- Langkah kedua adalah menentukan titik knot. Dalam penentuan titik knot, boleh mengambil sebarang titik dari data. Misalkan titik knot yang dipilih adalah 1 titik knot dari variabel x_1 dan 1 titik knot dari variabel x_2 yaitu $K_1 = 0,46$ dan $K_2 = 22,62$
- Membuat matriks kolom identitas (\mathbf{W}). Pada saat menentukan matriks kolom nilainya harus lebih besar sama dengan 0, atau dengan kata lain merupakan bilangan bulat positif. Karena data amatan sebanyak 10 maka akan terbentuk matriks berukuran 10×5 . Dengan rumus $[1, x_1, (x_1 - K_1), x_2, (x_2 - K_2)]$ diperoleh matriks \mathbf{W} sebagai berikut:

$$W = \begin{bmatrix} 1 & 1,18 & 0,69 & 42,44 & 19,82 \\ 1 & 0,39 & 0,00 & 25,19 & 2,57 \\ 1 & 0,55 & 0,06 & 18,90 & 0,00 \\ 1 & 0,33 & 0,00 & 22,99 & 0,37 \\ 1 & 0,60 & 0,11 & 27,51 & 4,89 \\ 1 & 0,59 & 0,10 & 24,21 & 1,59 \\ 1 & 0,72 & 0,23 & 22,62 & 0,00 \\ 1 & 1,04 & 0,55 & 23,99 & 1,37 \\ 1 & 0,49 & 0,00 & 28,33 & 5,71 \\ 1 & 0,95 & 0,49 & 52,74 & 30,12 \end{bmatrix}$$

4. Membuat transpose matriks W yang berukuran 5×10

$$W^T = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1,18 & 0,39 & 0,55 & 0,33 & 0,60 & 0,59 & 0,72 & 1,04 & 0,49 & 0,95 \\ 0,69 & 0,00 & 0,06 & 0,00 & 0,11 & 0,10 & 0,23 & 0,55 & 0,00 & 0,46 \\ 42,44 & 25,19 & 18,90 & 22,99 & 27,51 & 24,21 & 22,62 & 23,99 & 28,33 & 52,74 \\ 19,82 & 2,57 & 0,00 & 0,37 & 4,89 & 1,59 & 0,00 & 1,37 & 5,71 & 30,12 \end{bmatrix}$$

5. Menentukan nilai β dengan menggunakan metode *ordinary least square* (OLS)

menurut persamaan 2.28 dapat menggunakan rumus $\beta = (W^T * W)^I * W^T * Y$

$$W^T * W = \begin{bmatrix} 10,00 & 6,8400 & 2,2000 & 288,9200 & 66,4400 \\ 6,84 & 5,4066 & 2,1468 & 213,8956 & 61,2208 \\ 2,20 & 2,1468 & 1,0688 & 78,5222 & 28,9814 \\ 288,92 & 213,8956 & 78,5222 & 9335,6450 & 2870,5826 \\ 66,44 & 61,2208 & 28,9814 & 2870,5826 & 1367,7098 \end{bmatrix}$$

$$(W^T * W)^I = \begin{bmatrix} 67,147545 & -34,6825319 & 39,1157285 & -2,34594980 & 2,38546477 \\ -34,682532 & 47,2023758 & -52,1737689 & 0,62707150 & -0,63862338 \\ 39,115728 & -52,1737689 & 60,4602607 & -0,74326832 & 0,71408494 \\ -2,345950 & 0,6270715 & -0,7432683 & 0,09401266 & -0,09567459 \\ 2,385465 & -0,6386234 & 0,7140849 & -0,09567459 & 0,09910976 \end{bmatrix}$$

$$\text{Determinan} = 829.1663$$

Ketika matriks $(W^T * W)$ menghasilkan nilai determinan yang sama dengan 0 atau biasa disebut matriks singular, maka nilai GCV tidak akan dapat ditemukan. Hal ini disebabkan karena kita tidak dapat menentukan invers dari matriks $(W^T * W)$ yang nilai determinannya sama dengan 0.

$$(W^T * W)^I * W^T = \begin{bmatrix} & [1] & [2] & [3] & [4] \\ [1] & 0,92981252 & 0,65752679 & 6,080645e + 00 & 2,65154576 \\ [2] & -1,02822999 & -2,11893636 & -4,524381e - 12 & -4,92566477 \\ [3] & 1,87711718 & 1,88022796 & 5,037748e - 12 & 5,07485752 \\ [4] & -0,02523350 & 0,02090341 & -2,688172e - 01 & -0,01306463 \\ [5] & 0,02853359 & -0,01892925 & 2,688172e - 01 & 0,01183078 \end{bmatrix}$$

$$\begin{bmatrix} & [5] & [6] & [7] & [8] \\ [1] & -2,23140004 & -2,40613142 & -1,89264466 & -0,41988624 \\ [2] & 2,02764761 & 2,11548276 & 1,48756915 & -0,11910272 \\ [3] & -2,49334026 & -2,47990004 & -1,35625454 & 1,25544161 \\ [4] & 0,06697324 & 0,07361957 & 0,06115645 & 0,02169663 \\ [5] & -0,06652122 & -0,07861190 & -0,07426382 & -0,04540993 \end{bmatrix}$$

$$\begin{bmatrix} & [9] & [10] \\ [1] & -2,68664939 & 0,317181518 \\ [2] & 2,56502831 & -0,003793973 \\ [3] & -3,42878473 & -0,329364705 \\ [4] & 0,07839211 & -0,015626082 \\ [5] & -0,07200515 & 0,046559686 \end{bmatrix}$$

$$(W^T * W)^I * W^T * Y = \begin{bmatrix} -6,1063357 \\ 21,9964676 \\ -22,1334243 \\ 0,5950415 \\ -0,5064469 \end{bmatrix}$$

6. Menghitung nilai taksiran variabel respon (\hat{y}) serta residual yang diperoleh yaitu:

$$\hat{y} = W * ((W^T * W)^I * W^T * Y) = \begin{bmatrix} 19,79322 \\ 16,15981 \\ 15,91000 \\ 14,64512 \\ 18,54993 \\ 18,25894 \\ 18,10027 \\ 18,17782 \\ 18,63765 \\ 20,73724 \end{bmatrix}$$

$$Residual = Y - \hat{Y} = \begin{bmatrix} 23,82 \\ 15,72 \\ 15,91 \\ 14,92 \\ 18,38 \\ 15,84 \\ 18,28 \\ 16,09 \\ 22,34 \\ 17,67 \end{bmatrix} - \begin{bmatrix} 19,79322 \\ 16,15981 \\ 15,91000 \\ 14,64512 \\ 18,54993 \\ 18,25894 \\ 18,10027 \\ 18,17782 \\ 18,63765 \\ 20,73724 \end{bmatrix} = \begin{bmatrix} 4,026784e + 00 \\ -4,398130e - 01 \\ -2,160050e - 12 \\ 2,748831e - 01 \\ -1,699339e - 01 \\ -2,418941e + 00 \\ 1,797283e - 01 \\ -2,087820e + 00 \\ 3,702353e + 00 \\ -3,067240e + 00 \end{bmatrix}$$

7. Menentukan nilai *mean square error*. Menurut persamaan 2.31 menghitung nilai MSE dengan rumus

$$MSE = n^{-1}((I - G)Y)^T (I - G)Y$$

dengan $G = W(W^T * W)^{-1} * W^T$

didapat

$$G = \begin{bmatrix} & [1] & [2] & [3] & [4] \\ [1] & 5,063382e-01 & -3,349760e-02 & 1,232903e-13 & 2,093600e-02 \\ [2] & -3,349760e-02 & 3,090505e-01 & 1,762479e-13 & 4,318435e-01 \\ [3] & 2,415845e-13 & 6,246115e-13 & 1,000000e+00 & 1,381728e-12 \\ [4] & 2,093600e-02 & 4,318435e-01 & 4,428680e-13 & 7,300978e-01 \\ [5] & -3,528679e-02 & 7,547896e-02 & -2,129408e-13 & -4,717435e-02 \\ [6] & -5,466597e-02 & 7,135129e-02 & -2,231548e-13 & -4,459456e-02 \\ [7] & 5,044222e-02 & 3,718027e-02 & -1,465494e-13 & -2,323767e-02 \\ [8] & 3,266072e-01 & -3,650181e-02 & 5,606626e-15 & 2,281363e-02 \\ [9] & -1,259583e-01 & 1,033557e-01 & -2,624567e-13 & -6,459731e-02 \\ [10] & 3,450850e-01 & 4,173926e-02 & 1,587619e-14 & -2,608704e-02 \end{bmatrix}$$

$$\begin{bmatrix} & [5] & [6] & [7] & [8] \\ [1] & -3,528679e-02 & -5,466597e-02 & 5,044222e-02 & 3,266072e-01 \\ [2] & 7,547896e-02 & 7,135129e-02 & 3,718027e-02 & -3,650181e-02 \\ [3] & -5,015988e-13 & -5,196399e-13 & -3,606004e-13 & 3,902434e-14 \\ [4] & -4,717435e-02 & -4,459456e-02 & -2,323767e-02 & 2,281363e-02 \\ [5] & 2,280662e-01 & 2,312315e-01 & 1,699727e-01 & 2,157033e-02 \\ [6] & 2,312315e-01 & 2,513503e-01 & 2,119139e-01 & 8,846085e-02 \\ [7] & 1,699727e-01 & 2,119139e-01 & 2,498255e-01 & 2,738891e-01 \\ [8] & 2,157033e-02 & 8,846085e-02 & 2,738891e-01 & 6,050303e-01 \\ [9] & 2,796631e-01 & 2,672237e-01 & 1,447801e-01 & -1,228719e-01 \\ [10] & 7,647837e-02 & -2,227094e-02 & -1,147660e-01 & -1,789977e-01 \end{bmatrix}$$

$$\begin{bmatrix} & [9] & [10] \\ [1] & -1,259583e-01 & 3,450850e-01 \\ [2] & 1,033557e-01 & 4,173926e-02 \\ [3] & -6,366019e-13 & -2,664535e-14 \\ [4] & -6,459731e-02 & -2,608704e-02 \\ [5] & 2,796631e-01 & 7,647837e-02 \\ [6] & 2,672237e-01 & -2,227094e-02 \\ [7] & 1,447801e-01 & -1,147660e-01 \\ [8] & -1,228719e-01 & -1,789977e-01 \\ [9] & 3,799136e-01 & 1,384914e-01 \\ [10] & 1,384914e-01 & 7,403277e-01 \end{bmatrix}$$

$$(I - G) * Y = \begin{bmatrix} 4,026784e + 00 \\ -4,398130e - 01 \\ -7,950901e - 13 \\ 2,748831e - 01 \\ -1,699339e - 01 \\ -2,418941e + 00 \\ 1,797283e - 01 \\ -2,087820e + 00 \\ 3,702353e + 00 \\ -3,067240e + 00 \end{bmatrix}$$

$$((I - G) * Y)^T = \begin{bmatrix} [1] & [2] & [3] \\ [1] & 4,026784e + 00 & -4,398130e - 01 & -7,950901e - 13 \end{bmatrix}$$

$$\begin{bmatrix} [4] & [5] & [6] & [7] & [8] \\ 2,748831e - 01 & -0,1699339 & -2,418941 & 0,1797283 & -2,08782 \end{bmatrix}$$

$$\begin{bmatrix} [9] & [10] \\ [1] & 3,702353 & -3,06724 \end{bmatrix}$$

$$((I - G) * Y)^T * ((I - G) * Y) = 49,87082$$

$$MSE = \frac{((I - G) * Y)^T * ((I - G) * Y)}{n} = \frac{49,87082}{10} = 4,987082$$

8. Menentukan nilai *generalized cross validation* (GCV) dengan persamaan 2.34

didapat:

$$Tr(I - G) = 5$$

$$\frac{Tr(I - G)}{n} = \frac{5}{10} = 0,5$$

$$\left(\frac{Tr(I - G)}{n}\right)^2 = (0,5)^2 = 0,25$$

$$GCV = \frac{MSE}{(n^{-1}Tr(I - G))^2} = \frac{4,987082}{0,25} = 19,94833$$

2.13 Tingkat Kriminalitas

Kriminalitas merupakan segala macam bentuk tindakan dan perbuatan yang merugikan secara ekonomis dan psikologis yang melanggar hukum yang berlaku dalam negara Indonesia serta norma-norma sosial dan agama. Dapat diartikan bahwa, tindak kriminalitas adalah segala sesuatu perbuatan yang melanggar hukum dan melanggar norma-norma sosial, sehingga masyarakat menentangnya. (Kartono, 1999: 122)

Tindak kejahatan atau kriminalitas merupakan perbuatan seseorang yang dapat diancam hukuman berdasarkan KUHP atau Undang-Undang serta peraturan lainnya yang berlaku di Indonesia. Kriminalitas juga merupakan suatu permasalahan yang banyak dihadapi oleh negara. Kesuksesan pembangunan yang dibangun oleh setiap negara sangat bergantung terhadap besar kecilnya hambatan dari kriminalitas. Peran aktif dan dukungan masyarakat terhadap proses pembangunan akan optimal jika kriminalitas bisa ditekan serendah-rendahnya. Berbagai kerugian telah banyak ditimbulkan oleh adanya tindak kriminal, baik itu kerugian ekonomi, fisik, moral dan psikologis. Dari sudut pandang ekonomi, kriminalitas menimbulkan kerugian dengan adanya biaya yang harus dikeluarkan akibat dari tindak kriminal. Biaya tersebut tidak hanya ditanggung oleh korban, namun oleh masyarakat, dunia usaha, dan juga negara atau pemerintah daerah.

Menurut Badan Pusat Statistik, jenis kejahatan yang termasuk dalam tingkat kriminalitas antara lain kejahatan terhadap nyawa (pembunuhan), kejahatan terhadap

fisik (penganiayaan berat, penganiayaan ringan, dan kekerasan dalam rumah tangga), kejahatan terhadap kesusilaan (pemerkosaan dan pencabulan), kejahatan terhadap kemerdekaan orang (penculikan dan mempekerjakan anak dibawah umur), kejahatan terhadap hak milik/barang dengan penggunaan kekerasan (pencurian dengan kekerasan, pencurian dengan kekerasan menggunakan senjata api, dan pencurian dengan kekerasan dengan menggunakan senjata tajam), kejahatan terhadap hak milik/barang (pencurian, pencurian dengan pemberatan, pencurian kendaraan bermotor, penghancuran barang, pembakaran dengan senjata, dan penadahan), kejahatan terkait narkoba (narkoba dan psikotropika), kejahatan terkait penipuan, penggelapan dan korupsi (penipuan/perbuatan curang, penggelapan, dan korupsi), serta kejahatan terhadap ketertiban umum.

Berdasarkan ilmu kriminologi, kecenderungan individu untuk melakukan tindak kriminalitas dapat dilihat dari perspektif biologis, perspektif sosiologis, dan perspektif lainnya. Ilmu ini juga memberikan dua arti untuk istilah kejahatan, yakni secara yuridis dan sosiologis. Secara yuridis, Bonger berpendapat bahwa kejahatan berarti perbuatan anti sosial yang secara sadar mendapat reaksi dari negara berupa pemberian derita dan kemudian sebagai reaksi terhadap rumusan-rumusan hukum atau legal definitions mengenai kejahatan. Secara sosiologis, kejahatan merupakan suatu perilaku manusia yang diciptakan oleh masyarakat. (Rohman, 2016)

Ada dua faktor yang dapat menimbulkan kejahatan yaitu faktor intern yang meliputi sifat khusus dan sifat umum dalam diri individu, dan faktor ekstern. Sifat khusus dalam diri individu antara lain; sakit jiwa, daya emosional, rendahnya mental,

dan anatomi, sedangkan sifat umum dalam diri individu antara lain; umur, kekuatan fisik, kedudukan individu di dalam masyarakat, pendidikan individu, dan hiburan individu. Faktor ekstern dapat mencakup faktor-faktor ekonomi (perubahan harga, pengangguran, urbanisasi), faktor agama, faktor bacaan, dan faktor film. Pada umumnya para pelaku tindak kejahatan melakukan hal ilegal karena perkiraan kepuasan yang akan mereka dapatkan jauh lebih besar dibandingkan kepuasan yang pasti mereka dapatkan apabila mengikuti hukum yang berlaku atau perbuatan legal. (Rohman, 2016)

Menurut Todaro (2004) bahwa permintaan akan pendidikan dipengaruhi oleh dua hal, yaitu harapan bagi seorang siswa yang lebih terdidik untuk mendapatkan pekerjaan dengan hasil yang lebih baik pada sektor modern di masa yang akan datang bagi siswa itu sendiri maupun keluarganya serta biaya-biaya pendidikan baik yang bersifat langsung maupun tidak langsung yang harus dikeluarkan atau ditanggung oleh siswa dan/ keluarganya. Sedangkan dari sisi penawaran, jumlah sekolah di tingkat sekolah dasar, menengah, dan universitas lebih banyak ditemukan oleh proses politik, yang seringkali tidak berkaitan dengan kriteria ekonomi.

Teori Becker (1968) menjelaskan bahwa tingkat kesenjangan antara orang dengan pendapatan rendah dari sektor legal dibandingkan dengan orang yang berharta tinggi dalam suatu daerah. Hal ini menjelaskan bahwa ketimpangan distribusi pendapatan merupakan salah satu permasalahan dalam negara-negara berkembang. Distribusi pendapatan perseorangan sendiri merupakan ukuran yang paling sering digunakan oleh para ekonom untuk menghitung jumlah penghasilan yang diterima

oleh setiap individu atau rumah tangga (Todaro & Sith, 2004). Hal tersebut akan mengakibatkan aksi kejahatan akan naik di daerah tersebut akibat adanya *return* dari aksi kejahatan.

Secara umum, perbuatan kriminal dilakukan karena merasa belum terpenuhi kebutuhan hidup mereka selama ini. Kesenjangan sosial, kesejahteraan yang tak seimbang, instabilitas ekonomi juga menjadi beberapa pemicu lahirnya tindakan kejahatan (Prayetno, 2013). Oleh karena itu, peran pemerintah dan masyarakat sangat diperlukan untuk mengurangi jumlah tindak kejahatan.

Mayoritas remaja pelaku tindak pidana atau remaja nakal berasal dari keluarga yang kurang atau tidak mampu secara ekonomi. Fakta ini mendukung pendapat Marton (dalam Rutter dkk, 1998) yang mengatakan bahwa kemiskinan merupakan faktor yang penyebab timbulnya kenakalan pada remaja. Kenakalan tersebut muncul dari ketegangan yang diakibatkan adanya kesenjangan (*gap*) antara tujuan dengan sarana yang tersedia untuk mencapai tujuan tersebut. Kenakalan atau kriminalitas yang paling menonjol di kalangan remaja adalah berupa tindak pidana pencurian dan tindak pidana narkoba. Motivasi atau alasan remaja melakukan kejahatan tersebut nampak terkait dengan latar belakang keluarga dan latar belakang sosial mereka. Sebagian besar remaja melakukan tindak pidana pencurian dengan alasan faktor ekonomi, sedangkan alasan remaja melakukan tindak pidana narkoba pada umumnya adalah karena pengaruh atau ajakan dari teman sebayanya

Menurut BPS (2010) bahwa persentase remaja pelaku tindak pidana yang pada saat melakukan tindak pidana berstatus sebagai pelajar atau masih sekolah hanya

sebesar 38,0 persen. Persentase tersebut untuk kelompok remaja yang tidak bersekolah lagi mencapai hampir dua kali lipatnya atau sebesar 60,0 persen. Dari jumlah keseluruhan remaja tersebut, sebesar 2,0 persen remaja tidak pernah bersekolah. Pendidikan tertinggi yang sedang atau pernah dijalani oleh para remaja pelaku tindak pidana pada saat mereka melakukan perbuatan kriminalitas mayoritas adalah belum/tidak tamat SMTP (34,5 persen), kemudian belum/tidak tamat SD (27,5 persen) dan belum/tidak tamat SMTA (17,5 persen).

Proses pendidikan akan melahirkan sumberdaya manusia yang memiliki kemampuan dan keterampilan sebagai bekal untuk berperan dan berinteraksi dalam masyarakat. Analisis yang dilakukan oleh Ehrlich (1996) yaitu pendidikan merupakan suatu hal yang dibutuhkan untuk mencegah kejahatan, karena dengan pendidikan yang tinggi seseorang akan memperhitungkan resiko apabila melakukan suatu tindak kejahatan. Dengan alasan tersebut pendidikan akan mampu menekan angka tindak kejahatan.

Berdasarkan artikel Lampost.co, salah satu faktor yang menyebabkan peningkatan tindak kejahatan yang tidak diikuti oleh peningkatan kemiskinan adalah kemampuan kepolisian dalam menyelesaikan tindak kriminal. Berdasarkan data terbaru yang tersedia, persentase penyelesaian tindak pidana oleh kepolisian meningkat sebesar 8,79% dari angka 58,9% di tahun 2014 menjadi 67,69% di tahun 2015. Menurut teori kejahatan yang merupakan proses pengambilan keputusan yang rasional, maka dengan terjadi peningkatan kinerja polisi dalam mengungkap kasus kejahatan

membuat pertimbangan tersebut seakan menjadi seimbang. Kejahatan yang terjadi bisa saja terjadi lebih banyak. Namun, hanya terjadi sedikit peningkatan karena masyarakat miskin yang semula berniat melakukan tindak kejahatan mengurungkan niatnya. Risiko tertangkap yang lebih besar membuat mereka berpikir dua kali dan lebih memilih untuk mencari sumber pemasukan lain. Di sinilah sumbangsih penting kepolisian dalam memutus hubungan antara peningkatan tindak kejahatan dan peningkatan kemiskinan

2.14 Penelitian yang Relevan

Penelitian yang relevan dengan penelitian ini adalah sebagai berikut.

1. Ratnasari (2015) dengan penelitiannya yang membandingkan performansi regresi *spline* dan deret fourier untuk regresi nonparametrik multivariabel pada data kemiskinan di Papua. Nilai minimum GCV pada *spline truncated* (16,70) lebih kecil dari deret fourier (18,79). Nilai R^2 pada *spline* (98,46%) lebih besar dari deret fourier (89,20%). Nilai MSE dari *spline* (4,61) lebih kecil dari deret fourier (8,94). Jadi dapat disimpulkan bahwa metode *Spline* lebih baik dari pada metode deret fourier.
2. Aydin (2007) tentang perbandingan teknik yang digunakan untuk memprediksi regresi nonparametrik model dengan melakukan *smoothing* estimator regresi *spline* dan regresi kernel. Nilai MSE yang diperoleh adalah *spline* (174123,8) dan kernel (266454,99), Nilai RMSE sebesar 87061,9 untuk *spline* dan

133227,4 untuk kernel, Nilai MAE sebesar 315,0 dan 378,8 untuk kernel, dan nilai MAPE sebesar 1679,4 untuk *spline* dan 1494,8 untuk kernel. Dilihat dari nilai MSE, RMSE, MAE, dan MAPE, model nonparametrik berdasarkan pemulusan *spline* memiliki performansi yang baik.

3. Astiti, Sumarjaya, dan Susilawati (2016) mengkaji tentang model regresi nonparametric spline multivariate terbaik yang menjelaskan hubungan antara variabel-variabel yang diduga berpengaruh terhadap indikator kemiskinan di Indonesia yang menghasilkan model regresi spline kubik dengan lima titik knot sebagai model terbaik dengan nilai GCV 0,140 dan nilai koefisien determinasi (R^2) sebesar 99,99%.
4. Sari (2017) menganalisis perbandingan model regresi nonparametric spline multivariabel dengan menggunakan metode *Generalized Cross Validation* (GCV) dan metode *Unbiassed Risk* (UBR) dalam pemilihan titik knot optimal. Pada hasil penelitian didapatkan menggunakan metode GCV menghasilkan nilai MSE 0,002059 dan pada uji parameter didapatkan semua variabel berpengaruh signifikan dan semua asumsi residual terpenuhi dengan nilai R_{adj}^2 adalah sebesar 92,4%. Sedangkan dengan metode UBR didapatkan nilai MSE sebesar 0,01315 dan pada uji parameter didapatkan bahwa semua variabel tidak berpengaruh signifikan dan asumsi residual distribusi normal tidak terpenuhi dengan nilai R_{adj}^2 adalah sebesar 52,15%. Hal ini membuktikan bahwa pemodelan regresi nonparametric dengan menggunakan metode GCV lebih

baik untuk data angka kematian maternal dibandingkan menggunakan metode UBR dalam pemilihan titik knot.

5. Marina (2013) tentang pemodelan regresi semiparametrik spline dan aplikasinya pada persentase kriminalitas di Jawa Timur yang menghasilkan faktor-faktor yang berpengaruh secara signifikan terhadap persentase kriminalitas di Jawa Timur antara lain kepadatan penduduk, tingkat pengangguran terbuka, persentase penduduk miskin, persentase penduduk yang tidak pernah sekolah, persentase korban penyalahgunaan NAPZA, dan persentase keluarga bermasalah.
6. Dermawanti, Abdul, dan Rusgiyono (2015) menganalisis tentang faktor-faktor yang mempengaruhi kriminalitas di Kabupaten Batang tahun 2013 dengan analisis jalur menghasilkan faktor yang memiliki pengaruh positif terbesar terhadap kriminalitas adalah faktor pengangguran dengan pengaruh langsung sebesar 0,395. Faktor yang memiliki pengaruh positif terbesar kedua adalah faktor pendidikan dengan pengaruh langsung sebesar 0,222 dan pengaruh tidak langsung 0,0818. Sedangkan, faktor yang memiliki pengaruh positif terkecil adalah faktor moral dengan pengaruh sebesar 0,180.
7. Dona dan Setiawan (2015) tentang pemodelan faktor-faktor yang mempengaruhi tingkat kriminalitas di Jawa Timur dengan analisis regresi spasial menghasilkan Berdasarkan metode SEM, ada 2 variabel yang berpengaruh signifikan terhadap tingkat kriminalitas, yaitu variabel kepadatan penduduk yang berpengaruh positif dan persentase penduduk miskin yang

berpengaruh negatif. Sedangkan dengan menggunakan metode SAR, ada 3 variabel yang berpengaruh signifikan terhadap tingkat kriminalitas di Jawa Timur yaitu kepadatan penduduk dan Indeks gini yang berpengaruh positif serta PDRB perkapita yang berpengaruh negatif.

2.15 Kerangka Berpikir

Analisis regresi merupakan suatu metode yang digunakan untuk menganalisis hubungan antara dua atau lebih variabel. Menurut Hardle (1990:4) untuk mengestimasi fungsi regresi ada dua pendekatan yang dapat digunakan dalam menentukan kurva regresi yaitu pendekatan parametrik dan pendekatan regresi nonparametrik. Pendekatan yang sering kali digunakan adalah pendekatan regresi parametrik yang fungsi regresinya diketahui berdasarkan pengalaman, informasi sebelumnya atau bentuk kurva regresinya diketahui. Sedangkan pendekatan model regresi nonparametrik digunakan berdasarkan pendekatan yang tidak terikat dengan asumsi bentuk kurva tertentu dan memberikan fleksibilitas yang lebih besar. Menurut Eubank (1999:9) model regresi nonparametrik mempunyai kelebihan yaitu fleksibilitas yang besar untuk menyesuaikan dengan data aslinya daripada model regresi parametrik.

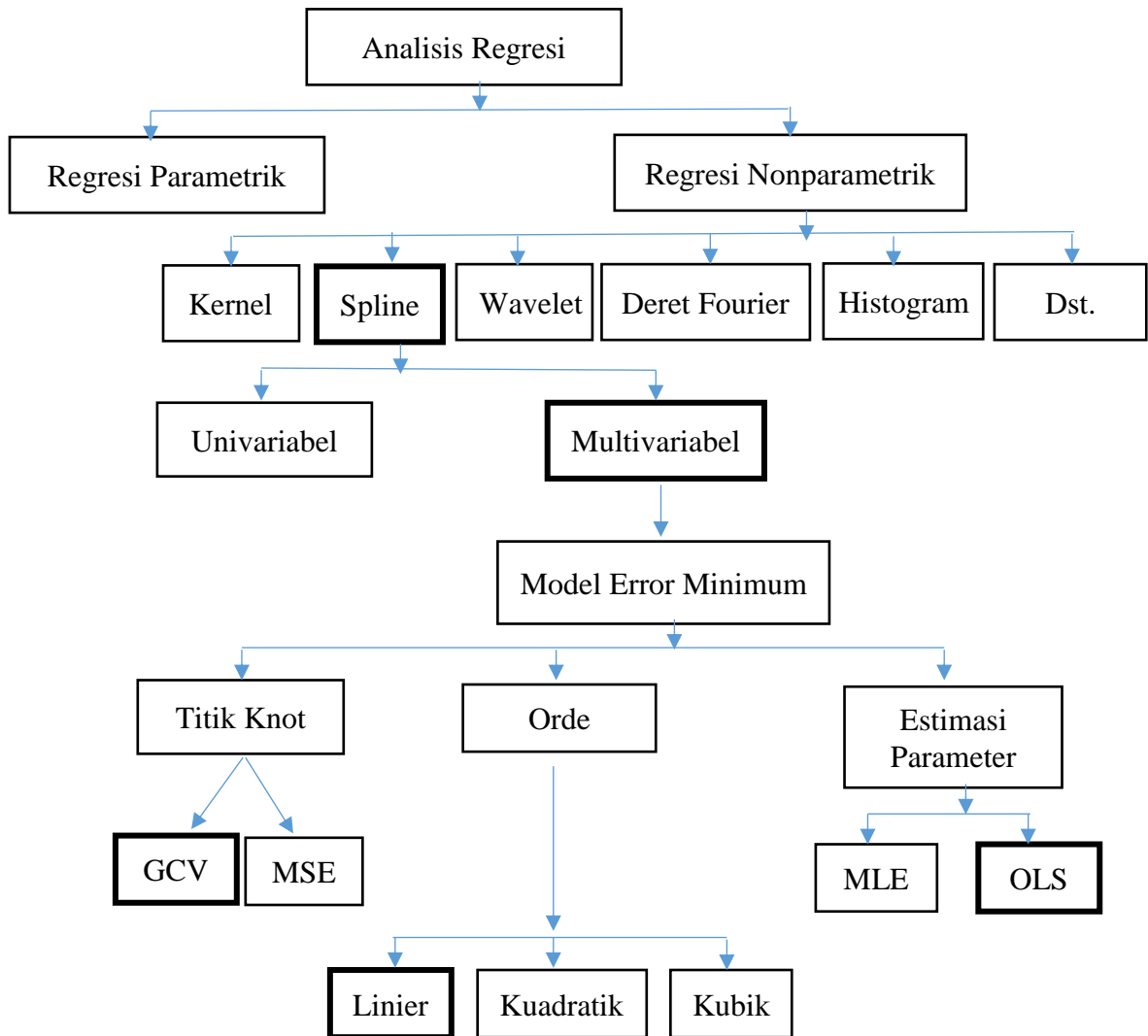
Beberapa teknik estimasi dalam regresi nonparametrik yaitu pendekatan histogram, estimator *spline*, estimator *kernel*, estimator orthogonal, analisis wavelet dan lain-lain. Salah satu yang banyak dikaji dalam regresi nonparametrik adalah regresi *spline*. Regresi *spline* adalah suatu pendekatan kearah plot data dengan tetap

memperhitungkan kemulusan kurva. Dilihat dari banyaknya variabel prediktor dalam regresi dapat dibedakan menjadi dua, yaitu univariabel dan multivariabel. Univariabel hanya mempunyai satu variabel prediktor sedangkan multivariabel mempunyai lebih dari satu variabel prediktor untuk diteliti. Dalam regresi *spline*, apabila terdapat satu variabel respon dan prediktor maka disebut dengan regresi nonparametrik *spline* univariabel dan apabila terdapat lebih dari satu variabel respon dan prediktor maka disebut dengan regresi nonparametrik *spline* multivariabel.

Penggunaan regresi *spline* sangat bergantung pada pemilihan titik knot dan orde. Salah satu metode yang dapat digunakan dalam penentuan titik knot adalah dengan mencari nilai GCV yang paling minimum. Pada penentuan orde, ketika terjadi perubahan pola data naik dan turun secara tajam dapat menggunakan orde linier, ketika terjadi perubahan pola data naik secara halus dapat menggunakan orde kuadrat, dan apabila kurva yang ditunjukkan terjadi perubahan pola data naik dan turun secara halus dapat menggunakan orde kubik. Perubahan perilaku pola data seperti ini yang dapat digambarkan dengan sangat baik dengan menggunakan pendekatan regresi *spline*.

Beberapa metode yang dapat digunakan dalam pemilihan estimasi parameter diantaranya metode MLE (*Maximum Likelihood Error*) dan OLS (*Ordinary Least Square*). Metode *maximum likelihood estimation* (MLE) bekerja dengan memaksimumkan fungsi *likelihood* $L(\theta)$ atau fungsi *log-likelihood* $\ln L(\theta)$. Selain *maximum likelihood estimation* (MLE), metode yang biasa digunakan untuk menaksir

parameter adalah *ordinary least square (OLS)*. Dalam penelitian ini menggunakan metode OLS karena pengerjaannya lebih mudah.



Gambar 2.1 Kerangka Berpikir

BAB 5

PENUTUP

5.1 Simpulan

Berdasarkan rumusan masalah dari hasil pembahasan maka dapat diambil kesimpulan sebagai berikut

1. Model regresi spline multivariable terbaik pada data factor yang mempengaruhi tingkat kriminalitas di Jawa Tengah tahun 2016 berdasarkan nilai GCV minimum yaitu model regresi spline linier dengan 8 titik knot dan 2 variabel yaitu

$$\begin{aligned}\hat{Y}(x) = & -0,4814473 + 0,0541275x_2 + 13,1338065(x_2 - 13,49)_+ \\ & - 39,7470132(x_2 - 13,59)_+ + 40,4197012(x_2 - 13,69)_+ \\ & - 14,1241691(x_2 - 13,89)_+ + 0,1341433x_5 + \\ & - 2,1887713(x_5 - 8,34)_+ + 2,0512095(x_5 - 8,84)_+ \\ & - 7,8266389(x_5 - 16,84)_+ + 15,0780795(x_5 - 17,84)_+\end{aligned}$$

2. Variabel yang berpengaruh terhadap tingkat kriminalitas adalah harapan lama sekolah (X_2) berpengaruh sebesar 87,84985%, persentase penduduk berumur 15 tahun ketas yang belum pernah sekolah (X_3) berpengaruh 42,024804%, dan persentase penduduk berumur 15 tahun keatas yang tamat SMP (X_5) berpengaruh sebesar 60,926040%.

3. Variabel yang signifikan terhadap model yaitu harapan lama sekolah (X_2) dan persentase penduduk berumur 15 tahun keatas yang tamat SMP (X_5) karena model regresi spline tersebut menghasilkan koefisien determinasi sebesar 97,97%,

5.2 Saran

Adapun saran yang diberikan peneliti untuk penelitian ini adalah

1. Hubungan antara tingkat kriminalitas dengan harapan lama sekolah dan penduduk berusia 15 tahun ke atas yang tamat SMP termasuk kategori sangat kuat, maka harapan lama sekolah dan penduduk berusia 15 tahun ke atas yang tamat SMP perlu ditekan agar tingkat kriminalitas berkurang. Serta sekolah juga memiliki peran penting dalam membentuk karakter yang baik. Penambahan mata pelajaran seperti pendidikan karakter dan budi pekerti sangat penting untuk membentuk anak dengan kepribadian yang baik karena usia remaja merupakan masa rawan akan tindak kejahatan.
2. Pada penelitian ini hanya terbatas pada orde linier, untuk penelitian selanjutnya dapat menambahkan orde seperti orde kuadratik dan kubik pada model regresi nonparametrik *spline* dengan kombinasi titik knot sehingga memungkinkan mendapatkan hasil yang lebih baik.
3. Penelitian ini hanya menggunakan 5 variabel. Pada penelitian selanjutnya diharapkan menambah banyaknya variabel pada regresi *spline* multivariabel.

DAFTAR PUSTAKA

- Adisasmita, Rahardjo. (2011). *Pengelolaan Pendapatan & Anggaran Daerah*. Yogyakarta: Graha Ilmu.
- Alizah, D. N., Dwidayati, N. K., Sugiman. (2017). *Estimasi Regresi Nonparametrik Penalized Spline Filter Menggunakan Metode Penalized Least Square Pada Data Time Series*. Unnes Journal of Mathematics Vol.6(2).
- Anderson, B. 2005. *Regression III: Advance Methods, Lecturer 15: regression spline*.
- Astiti, Sumarjaya, dan Susilawati. (2016). *Analisis Regresi Nonparametrik Spline Multivariat Untuk Pemodelan Indikator Kemiskinan di Indonesia*. E-Jurnal Matematika Vol. 5 No. 3. Jurusan Matematika Universitas Udayana.
- Astuti, N. R. (2014). [Skripsi]. Analisis Tingkat Kriminalitas di Kota Semarang Dengan Pendekatan Ekonomi Tahun 2010-2012. Fakultas Ekonomika dan Bisnis Universitas Diponegoro; Semarang.
- Aydin. (2007). *A comparison of the Nonparametric Regresion Models Using Smothing Spline and Kernel Regression*. World Academy of Science, Engineering and Technology 36. Munjla Universitesi.
- Becker, Gary. 1968. *Crime and Punishmen : An Economics Approach*. Journal of Political Economy.
- Bintianingrum dan Budiantara, I.N. (2014). *Pemodelan Regresi Nonparametrik Spline Truncated dan Aplikasinya pada Angka Kelahiran Kasar di Surabaya*. Jurnal Sains dan Seni POMITS Vo. 3 No. 1. Institut Teknologi Sepuluh Nopember.
- Budiantara, I.N. (2004). *Konsistensi Estimator Spline Terbobot Berdasarkan Kriteria Integrated Mean Square Error*. Prosiding Konferensi Nasional Matematika, XII Himpunan Matematika Indonesia Jurusan Matematika FMIPA UNUD Bali, ISBN : 979-99592-0-9.
- Budiantara, I.N. (2006). *Regresi Nonparametrik Dalam Statistika*. Makalah Pembicara Utama pada Seminar Nasional Matematika, Jurusan Matematika, Jurusan Matematika, FMIPA, Universitas Negeri Makassar (UNM), Makasar.

- Budiantara, I.N. (2007). *Pendugaan Model Fertilitas Wanita di Indonesia dengan Menggunakan Regresi Spline*. Laporan Akhir Pelaksanaan Penelitian Studi Kajian Wanita Tahun Anggaran 2007, LPPM Teknologi Sepuluh Nopember Surabaya.
- Budiantara, I.N. (2009). *Penelitian Bidang Regresi Spline Menuju Terwujudnya Penelitian Statistika yang Mandiri dan Berkarakter*. Seminar Nasional FMIPA Undiksha.
- [BPS] Badan Pusat Statistik. 2010. *Kriminalitas Remaja 2010*. Jakarta (ID): BPS.
- [BPS] Badan Pusat Statistik. 2016. *Statistik Kriminal 2016*. Jakarta (ID): BPS.
- Conover, W. J. (1999). *Practical Nonparametric Statistics Third Edition*. John Willey And Sons, Inc.
- Dermawanti, Hoyyi. A, dan Rusgiyono, A. (2015). *Faktor-Faktor yang Mempengaruhi Kriminalitas di Kabupaten Batang Tahun 2013 Dengan Analisis Jalur*. Jurnal Gaussian Vol.4(2).
- Dona dan Setiawan. (2015). *Pemodelan Faktor-Faktor Yang Mempengaruhi Tingkat Kriminalitas di Jawa Timur dengan Analisis Regresi Spasial*. Jurnal Sains dan Seni ITS Vol.4(1).
- Efromovich, Sam. (1999). *Nonparametric Curve Estimation: Method, Theory, And Applications*. Springer Series in Statistics.
- Ehrlich, Isaach, 1996. *Crime, Punishmen and Market for Offense*. The Journal of Economic Perspektive, Vol. 10(1).
- Eubank, R. (1988). *Spline Smoothing and Nonparametric Regression*. New York: Marcel Dekker.
- Eubank, R. (1999). *Nonparametric Regression and Smoothing Spline*, Marcel Dekker Inc.
- Gujarati, D. N. (2003). *Basic Econometrics Fourth Edition*. The McGraw-Hili Companies, Inc. 1221 Avenue of the Americas, New York.
- Hardle, W. (1990). *Applied Nonparametric Regression*. New York: Cambridge University Press.

- Kartono. 1999. *Patologi Sosial*. Jakarta: Raja grafindo Persada.
- Kutner, M. H., Nachtsheim, C. J., Neter, J, and Li, W. (2005). *Aplied Linear Statistical Models. Fifth Edition*. New York: McGraw-Hill.
- Lochner, Lance. 2007. *Education and Crime*. University of Westerb Ontario
- Mankiw, N. Gregory. (2007). *Macroeconomi : Edisi Keenam*. Erlangga : Jakarta.
- Marina, S. M. T. dan Budiantara, I. N. (2013). *Pemodelan Faktor-Faktor yang Mempengaruhi Persentase Kriminalitas di Jawa Timur dengan Pendekatan Regresi Semiparametrik Spline*. Jurnal Sains Dan Seni POMITS Vol. 2(2).
- Montgomery, D. C. (2013). *Design and Analysis of Experiments. Eighth Edition*. New York: John Wiley & Sons, Inc.
- Prayetno. 2013. *Kausalitas Kemiskinan Terhadap Perbuatan Kriminal (Pencurian)*. Media Komunikasi FIS Vol. 12(1).
- Ratnasari, V., I Nyoman B., dan Ismani Z. (2015). *Comparison Truncated Spline and Fourier Series in Multivariable Nonparametrik Regression Models (Applications: Data Of Poverty in Papua, Indonesia)*. International Journal of Basic & Applied Sciences Vol. 15 No. 04. Statistics Departement, Institut Teknologi Sepuluh November, Surabaya.
- Rohman, A. 2016. *Upaya Menekan Angka Kriminalitas Dalam Meretas Kejahatan Yang Terjadi Pada Masyarakat*. Jurnal Perspektif Vol XXI (2) Fakultas Hukum Universitas Borneo tarakan.
- Sari, S. U. R. (2017). *Perbandingan Model Regresi Nonparametrik Spline Multivariabel dengan Menggunakan Metode Generalized Cross Validation (GCV) dan Metode Unbiased Risk (UBR) dalam Pemilihan Titik Knot Optimal*. Prosiding Seminar Nasional Integrasi Matematika dan Nilai Islami Vol. 1 No.1. Fakultas Tarbiyah dan Ilmu Keguruan. Universitas Islam Negeri Maulana Malik Ibrahim Malang.
- Soleh, Achmad S. (2005). *Ilmu Statistika*. Penerbit Rekayasa Sains Bandung; Bandung.
- Suhartono. (2008). *Analisis Data Statistik dengan R. Lab. Statistik Komputasi, ITS, Surabaya*.
- Todaro, Michele P., and Stephen C. Smith Alih Bahasa. *Pembangunan Ekonomi*.(2016).

- Tripena, A. (2011). *Penentuan Model Regresi Spline Terbaik* . Prosiding Seminar Nasional Statistika Universitas Diponegoro Semarang, 92-102.
- Vitaningrum, R., Dwidayati, N. K., Sunarmi. (2018). *Analisis Regresi Nonparametrik dengan Fungsi Kernel*. Unnes Journal of Mathematics Vol.7(2)
- Wahba, G. (1990). *Spline Models For Observational Data*. SIAM, CBMS-NSF Regional Conference Series and Applied Mathematics.
- Wand, M.P. (2000). *A comparison of regression spline smoothing procedures*. *Computational Statistics* 15:443-62
- Wasserman, Larry. (2006). *All of Nonparametric Statistics*. Departement of Statistics Carnegie Mellon University. Pittsburgh: USA.