



**PENGEMBANGAN ANALISIS MODEL ADITIF TERGENERALISASI
DENGAN PENGHALUS *SPLINE* BERBASIS WEB INTERAKTIF
MENGUNAKAN R-SHINY**

SKRIPSI

Oleh
Dwindah Setiari Wigati
NIM 121810101054

**JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS JEMBER
2016**



**PENGEMBANGAN ANALISIS MODEL ADITIF TERGENERALISASI
DENGAN PENGHALUS *SPLINE* BERBASIS WEB INTERAKTIF
MENGUNAKAN R-SHINY**

SKRIPSI

diajukan guna melengkapi tugas akhir dan memenuhi salah satu syarat
untuk menyelesaikan Program Studi Matematika (S1)
dan mencapai gelar Sarjana Sains

Oleh
Dwindah Setiari Wigati
NIM 121810101054

**JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS JEMBER
2016**

PERSEMBAHAN

Skripsi ini saya persembahkan untuk:

1. ibunda Marokhaeni dan ayahanda Safi'i yang telah memberikan do'a, cinta dan semangat;
2. kakak tercinta Eri Ferdianto dan seluruh keluarga yang telah memberikan dukungan;
3. seluruh guru dan dosen sejak taman kanak-kanak hingga perguruan tinggi yang telah membimbing keingintahuan serta memberikan ilmu kepada saya dengan penuh perhatian dan kasih sayang;
4. almamater Jurusan Matematika FMIPA Universitas Jember, SMA Negeri 1 Mojosari, SMP Negeri 1 Ngoro, SDN 2 Jasem, serta TK Dharma Wanita.

MOTTO

“Maka sesungguhnya bersama kesulitan ada kemudahan. Sesungguhnya bersama kesulitan ada kemudahan. Maka apabila engkau telah selesai (dari suatu urusan), tetaplah bekerja keras (untuk urusan lain). Dan hanya kepada Tuhanmulah engkau berharap.”

(Q.S. Al-Insyirah, 6-8)¹

“Learn from yesterday, live for today, hope for tomorrow. The important thing is not to stop questioning.”

(Albert Einstein)²

¹ Departemen Agama Republik Indonesia. 1997. Al-Qur'an – Al-Karim dan Terjemahannya. Solo: CV.Pustaka Mantiq.

² http://www.brainyquote.com/quotes/authors/a/albert_einstein.html

PERNYATAAN

Saya yang bertanda tangan di bawah ini:

Nama : Dwindah Setiari Wigati

NIM : 121810101054

menyatakan dengan sesungguhnya bahwa karya ilmiah yang berjudul “Pengembangan Analisis Model Aditif Tergeneralisasi dengan Penghalus *Spline* Berbasis Web Interaktif Menggunakan R-Shiny” adalah benar-benar hasil karya sendiri, kecuali kutipan yang sudah saya sebutkan sumbernya, belum pernah diajukan pada institusi manapun, dan bukan karya jiplakan. Saya bertanggung jawab atas keabsahan dan kebenaran isinya sesuai dengan sikap ilmiah yang harus dijunjung tinggi.

Demikian pernyataan ini saya buat dengan sebenar-benarnya, tanpa ada tekanan dan paksaan dari pihak manapun serta bersedia mendapat sanksi akademik jika ternyata di kemudian hari pernyataan ini tidak benar.

Jember, Juni 2016

Yang menyatakan,

Dwindah Setiari Wigati

NIM 121810101054

SKRIPSI

**PENGEMBANGAN ANALISIS MODEL ADITIF TERGENERALISASI
DENGAN PENGHALUS *SPLINE* BERBASIS WEB INTERAKTIF
MENGUNAKAN R-SHINY**

Oleh

Dwindah Setiari Wigati
NIM 121810101054

Dosen Pembimbing Utama : Prof. Drs. I Made Tirta, M.Sc., Ph.D.

Dosen Pembimbing Anggota : Dian Anggraeni, S.Si., M.Si.

PENGESAHAN

Skripsi yang berjudul “**Pengembangan Analisis Model Aditif Tergeneralisasi dengan Penghalus *Spline* Berbasis Web Interaktif Menggunakan R-Shiny**“ telah diuji dan disahkan pada:

Hari, tanggal :

Tempat : Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas
Jember

Tim Penguji,

Ketua,

Sekretaris,

Prof. Drs. I Made Tirta, M.Sc., Ph.D.
NIP 197407192000121001

Dian Anggraeni, S.Si., M.Si.
NIP 198202162006042002

Penguji I,

Penguji II,

Dr. Mohamad Fatekurohman, S.Si., M.Si.
NIP 196906061998031001

Prof. Drs. Kusno, DEA., Ph.D.
NIP 196101081986021001

Mengesahkan,

Dekan,

Drs. Sujito, Ph.D.
NIP 196102041987111001

RINGKASAN

Pengembangan Analisis Model Aditif Tergeneralisasi dengan Penghalus *Spline* Berbasis Web Interaktif Menggunakan R-Shiny; Dwindah Setiari Wigati 121810101054; 2016; 80 halaman; Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Jember.

Model aditif tergeneralisasi (*Generalized Additive Models* atau GAM) adalah salah satu model statistika yang mengembangkan model aditif ke dalam bentuk distribusi keluarga eksponensial. Model aditif tergeneralisasi juga dapat disebut sebagai model pengembangan dari model linier tergeneralisasi, dengan mengganti prediktor linier dengan prediktor aditif. Model aditif tergeneralisasi dapat dituliskan ke dalam bentuk model nonparametrik maupun semiparametrik yang merupakan gabungan dari model parametrik dan nonparametrik. Pada model aditif tergeneralisasi ini, fungsi atau komponen nonparametrik diestimasi oleh fungsi penghalus yaitu penghalus diagram pencar. Penghalus diagram pencar yang digunakan dalam penelitian ini adalah penghalus *spline*.

Pada penelitian ini, analisis model aditif tergeneralisasi dengan penghalus *spline* yang biasa dikerjakan dengan bantuan program R berbasis skrip dikembangkan menjadi program analisis berbasis web interaktif menggunakan R-Shiny. Tujuannya adalah untuk memudahkan seseorang yang kurang mengerti program R berbasis skrip dalam melakukan analisis data menggunakan model aditif tergeneralisasi dengan penghalus *spline*, yaitu hanya dengan memasukkan data pada program web secara *online* dan memilih menu-menu yang disediakan. Program analisis model aditif tergeneralisasi dengan penghalus *spline* berbasis web interaktif ini memiliki kemampuan/fitur yang hampir sama dengan program R berbasis skrip. Program web interaktif ini dilengkapi dengan ringkasan teori mengenai model aditif tergeneralisasi

dengan penghalus *spline* yang berguna sebagai dasar dalam melakukan analisis data dan untuk menambah wawasan pengguna tentang salah satu model statistika ini. Selain itu, program web interaktif ini juga berbasis tutorial yang memudahkan pengguna dalam melakukan analisis data.

Program web interaktif yang telah dibuat kemudian diaplikasikan untuk menganalisis suatu data. Pada penelitian ini, program web tersebut diaplikasikan pada data produktivitas padi di Indonesia tahun 2013, yang terdiri atas data produktivitas padi sawah sebagai variabel respon (y), dan data luas lahan (x_1), jumlah petani (x_2), jumlah curah hujan (x_3), penyinaran matahari (x_4), kelembaban (x_5), serta temperatur (x_6) sebagai variabel prediktor. Pada bagian eksplorasi data, didapatkan variabel prediktor yang termasuk komponen parametrik adalah luas lahan (x_1) dan jumlah petani (x_2), sedangkan yang termasuk dalam komponen nonparametrik adalah jumlah curah hujan (x_3), penyinaran matahari (x_4), kelembaban (x_5), dan temperatur (x_6). Selain itu juga didapatkan distribusi terbaik untuk pencocokan model yaitu distribusi normal. Basis *spline* yang digunakan dalam pencocokan model pada penelitian ini adalah *cyclic cubic regression spline*. Pencocokan model dilakukan secara berulang-ulang dengan mengganti-ganti nilai dimensi basis (k) yang merupakan dimensi basis ditambah 2 pada program web interaktif, hingga didapatkan model terbaik dengan variabel prediktor yang signifikan dan memiliki nilai AIC yang minimum. Dari hasil pencocokan model didapatkan model terbaik untuk data produktivitas padi di Indonesia yaitu:

$$Y = 40,359 + 8,552 \times 10^{-6}x_1 + s(x_5, k = 14) + s(x_6, k = 14)$$

Atau dapat dituliskan sebagai berikut

$$Y = 40,359 + 8,552 \times 10^{-6}(\text{luas. lahan}) + s(\text{kelembaban}, k = 14) \\ + s(\text{temperatur}, k = 14)$$

PRAKATA

Puji syukur kehadiran Tuhan Yang Maha Esa atas segala rahmat, taufik, serta hidayah-Nya, sehingga penulis dapat menyelesaikan penulisan tugas akhir yang berjudul “Pengembangan Analisis Model Aditif Tergeneralisasi dengan Penghalus *Spline* Berbasis Web Interaktif Menggunakan R-Shiny”. Skripsi ini disusun untuk memenuhi salah satu syarat untuk menyelesaikan pendidikan strata satu (S1) pada Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Jember.

Penyusunan tugas akhir ini tidak luput dari bantuan berbagai pihak. Oleh karena itu, penulis ingin mengucapkan terima kasih kepada:

1. Bapak Prof. Drs. I Made Tirta, M.Sc., Ph.D. selaku Dosen Pembimbing Utama dan Ibu Dian Anggraeni, S.Si., M.Si. selaku Dosen Pembimbing Anggota yang telah bersedia meluangkan waktu, memberikan saran serta bimbingannya dalam penulisan skripsi ini;
2. Bapak Dr. Mohamad Fatekurohman, S.Si., M.Si. selaku Dosen Penguji I dan Bapak Prof. Drs. Kusno, DEA., Ph.D. selaku Dosen Penguji II yang telah memberikan kritik serta saran pada penulisan skripsi ini;
3. seluruh staf pengajar jurusan Matematika Fakultas MIPA Universitas Jember yang telah memberikan ilmu serta bimbingannya sehingga penulis dapat menyelesaikan skripsi ini;
4. ibunda Marokhaeni dan ayahanda Safi'i tercinta yang senantiasa memberi do'a, semangat dan kasih sayang;
5. kakakku tercinta Eri Ferdianto serta seluruh keluarga yang telah memberikan do'a dan semangat;
6. sahabat-sahabatku Yuni, Ade, Wahyu, Rere, Anggun, Ikfi, Desi, dan Vivi yang telah meluangkan waktu serta memberikan semangat, keceriaan dan canda tawa selama masa studi;

7. keluarga besar BATHICS'12 dan KKN 90 (Diah, Annisa, Arul, Ivan), serta teman-teman kos (asri, ulan, nadya) yang telah memberikan dukungan dan motivasi;
8. semua pihak yang tidak dapat saya sebutkan satu per satu.

Skripsi ini juga tidak lepas dari kekurangan dan kesalahn baik isi maupun susunannya. Oleh karena itu, penulis menerima segala kritik dan saran dari semua pihak demu kesempurnaan skripsi ini. Penulis berharap semoga skripsi ini dapat memberi manfaat bagi pembaca dan kehidupan.

Jember, Juni 2016

Penulis

DAFTAR ISI

	Halaman
HALAMAN JUDUL	i
HALAMAN PERSEMBAHAN	ii
HALAMAN MOTTO	iii
HALAMAN PERNYATAAN	iv
HALAMAN PEMBIMBING	v
HALAMAN PENGESAHAN	vi
RINGKASAN	vii
PRAKATA	ix
DAFTAR ISI	xi
DAFTAR GAMBAR	xiv
DAFTAR TABEL	xv
BAB 1. PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah	4
1.4 Tujuan	4
1.5 Manfaat	4
BAB 2. TINJAUAN PUSTAKA	5
2.1 Model Linier	5
2.2 Model Linier Tergeneralisasi	5
2.3 Model Aditif	6
2.4 Model Aditif Tergeneralisasi	7
2.5 Penghalus <i>Spline</i>	8
2.5.1 Regresi <i>Spline</i> Kubik	10
2.5.2 Regresi <i>Spline</i> Kubik Siklik	12

2.6	P-IRLS	15
2.7	Pemilihan Parameter Penghalus	15
2.8	R-Shiny	17
BAB 3.	METODE PENELITIAN	20
3.1	Data	20
3.2	Langkah-langkah Penelitian	20
3.2.1	Tahap Penyusunan Web Interaktif	20
3.2.2	Tahap Aplikasi Program Web Interaktif	24
BAB 4.	HASIL DAN PEMBAHASAN	27
4.1	Hasil dan Pembahasan Fitur Web Interaktif	27
4.1.1	Penyusunan Web Interaktif	27
4.1.2	Tampilan Web Interaktif	33
4.2	Hasil dan Pembahasan Analisis Data Produktivitas Padi	42
4.2.1	Eksplorasi Data Produktivitas Padi	42
4.2.2	Pencocokan Model Menggunakan Model Aditif Tergeneralisasi dengan Penghalus <i>Spline</i>	46
BAB 5.	PENUTUP	53
5.1	Kesimpulan	53
5.2	Saran	54
DAFTAR PUSTAKA		55
LAMPIRAN		57

DAFTAR GAMBAR

	Halaman
2.1 Contoh pencocokan model	6
2.2 Ilustrasi basis regresi <i>spline</i> kubik	12
2.3 Ilustrasi basis regresi <i>spline</i> kubik siklik	14
2.4 Pencocokan model menggunakan model aditif tergeneralisasi dengan penghalus <i>spline</i>	17
2.5 Struktur komponen R-Shiny	18
4.1 Langkah-langkah penelitian	26
4.1 Tampilan awal	34
4.2 Menu pilihan dan tampilan data	35
4.3 Tampilan <i>scatterplot matrix</i>	36
4.4 Tampilan nilai koefisien determinasi	37
4.5 Tampilan pemilihan distribusi	37
4.6 Tampilan pencocokan model	38
4.7 Tampilan hasil pencocokan model	39
4.8 Menu pilihan luaran (<i>output</i>)	40
4.9 Tampilan <i>term plot</i>	40
4.10 Hubungan antar variabel pada data produktivitas padi	43
4.11 Histogram distribusi untuk data produktivitas padi	45
4.12 <i>Term plot</i> model	52

DAFTAR TABEL

	Halaman
2.1 Definisi matriks untuk regresi <i>spline</i> kubik	11
2.2 Definisi matriks untuk regresi <i>spline</i> kubik siklik	13
4.1 Nilai R^2 variabel prediktor pada data produktivitas padi	44
4.2 Nilai AIC untuk pemilihan distribusi	46
4.3 Hasil pencocokan model $y \sim x_1 + x_2 + s(x_3) + s(x_4) + s(x_5) + s(x_6)$	47
4.4 Uji signifikansi model $y \sim x_1 + x_2 + s(x_3) + s(x_4) + s(x_5) + s(x_6)$	48
4.5 Hasil pencocokan model baru I	49
4.6 Uji signifikansi model $y \sim x_1 + x_2 + s(x_4) + s(x_5) + s(x_6)$	49
4.7 Hasil pencocokan model baru II	50
4.8 Uji signifikansi model $y \sim x_1 + s(x_4) + s(x_5) + s(x_6)$	50
4.9 Uji signifikansi model $y \sim x_1 + s(x_5) + s(x_6)$	51

BAB 1. PENDAHULUAN

1.1 Latar Belakang

Statistika merupakan salah satu ilmu yang memuat banyak cabang ilmu di dalamnya, salah satu diantaranya adalah analisis regresi. Analisis regresi merupakan salah satu bidang statistika yang memainkan peran sangat penting. Analisis regresi umumnya digunakan untuk menyelidiki pola hubungan fungsional antara variabel respon (variabel dependen) dan variabel prediktor (variabel independen). Selain itu, analisis regresi juga bermanfaat untuk peramalan (*forecasting*). Dewasa ini terdapat 3 (tiga) jenis model pendekatan regresi yang banyak dikembangkan oleh para peneliti, yaitu pendekatan regresi parametrik, regresi nonparametrik, dan regresi semiparametrik. Dalam model pendekatan regresi parametrik terdapat asumsi yang sangat kaku dan kuat, yaitu bentuk kurva regresinya diketahui, misalnya linier, kuadratik, kubik, dan lain-lain. Berbeda dengan model pendekatan regresi parametrik, dalam regresi nonparametrik bentuk kurva regresi tidak diketahui dan hanya diasumsikan halus (*smooth*), sehingga model ini memiliki fleksibilitas yang tinggi. Model pendekatan regresi semiparametrik merupakan gabungan antara model parametrik dan model nonparametrik.

Salah satu bentuk model nonparametrik adalah model aditif, jika semua fungsi f_j pada prediktor aditif $\eta(X) = \sum_{j=1}^p f_j(X_j)$ merupakan fungsi penghalus s_j . Model aditif merupakan perluasan model linier yang dikembangkan untuk menangani data yang berdistribusi normal namun fungsi yang dimiliki tidak harus linier. Model aditif kemudian dikembangkan menjadi model aditif tergeneralisasi (*Generalized Additive Models*) untuk menangani kondisi data yang variabel responnya tidak harus berdistribusi normal dan hubungan dengan variabel prediktor juga tidak harus linier. Hal tersebut membuat model aditif tergeneralisasi menjadi lebih fleksibel jika dibandingkan dengan model aditif. Model aditif tergeneralisasi dapat pula disebut sebagai model pengembangan dari model linier tergeneralisasi karena model ini

mengembangkan model linier tergeneralisasi dengan cara yang sama dengan model aditif mengembangkan model linier. Model ini diasumsikan berdistribusi keluarga eksponensial dan memiliki fungsi *link*.

Pencocokan model aditif tergeneralisasi menggunakan metode penghalusan berupa penghalus diagram pencar (*scatterplot smoother*) untuk mengestimasi fungsi mulus, dan menggunakan algoritma *Penalized Iteratively Re-weighted Least Square* (P-IRLS) untuk mengestimasi model. Adapun metode penghalusan yang dapat memberikan suatu hasil analisis numerik yang lebih baik menurut Hastie (1990) adalah penghalus *spline*. Beberapa statistisi seperti Hardle, Eubank, Cox dan O'Sullivan menyebutkan kelebihan *spline* dibanding dengan metode penghalusan lain, diantaranya *spline* dapat mengatasi pola data yang menunjukkan naik/turun yang tajam dengan bantuan titik-titik knot serta kurva yang dihasilkan relatif mulus, *spline* dapat menyesuaikan diri secara efektif terhadap karakteristik lokal dari data, dan *spline* memiliki interpretasi statistik serta interpretasi visual yang sangat khusus dan sangat baik. Penelitian menggunakan model aditif tergeneralisasi dengan penghalus *spline* telah banyak diterapkan sebelumnya, diantaranya Handayani (2006) meneliti pengaruh kecepatan angin dan kelembaban pada tingkat evaporasi (penguapan), Premono (2013) menduga nilai ujian nasional pada SMA Negeri 1 Situbondo, dan Alimi (2015) menganalisis desain kasus tunggal.

Model aditif tergeneralisasi dengan penghalus *spline* biasa dikerjakan dengan bantuan program statistika. Salah satu program statistika yang sering digunakan adalah R. Namun, pengerjaan di R sebagian besar menggunakan pendekatan skrip atau CLI (*Command Line Interface*) yang tidak mudah untuk peneliti yang bukan berlatar belakang statistika maupun matematika. Seiring dengan berjalannya waktu, tim R-Studio meluncurkan dua program pendukung R. Salah satunya adalah R-Shiny yang merupakan *toolkit* untuk membuat interaksi atau komunikasi antara dokumen HTML dengan program R yang memungkinkan orang dapat membuat laman web interaktif, sehingga kemampuan R yang pada dasarnya bersifat CLI bisa diakses melalui menu web secara GUI (*Graphical User Interface*) (Tirta, 2015). Dengan

adanya web interaktif tersebut akan memudahkan seseorang yang kurang mengerti program statistika dalam melakukan analisis data, yaitu hanya dengan memasukkan data pada program web interaktif secara *online*. Penelitian berbasis web interaktif menggunakan R-Shiny pernah dilakukan sebelumnya, diantaranya Cahyani (2015) memodelkan logistik multinomial. Wijiasih (2015) menerapkan *Generalized Estimating Equation* (GEE) untuk respon multinomial berskala ordinal.

Sejauh ini belum ada penelitian berbasis web interaktif menggunakan R-Shiny untuk model aditif tergeneralisasi dengan penghalus *spline*. Berdasarkan banyaknya peneliti yang menggunakan model aditif tergeneralisasi dengan penghalus *spline* dan kemudahan yang ditawarkan oleh web interaktif menggunakan R-Shiny, penulis tertarik untuk mengembangkan model aditif tergeneralisasi dengan penghalus *spline* berbasis web interaktif menggunakan R-Shiny. Data yang digunakan dalam penelitian ini adalah data sekunder dengan produktivitas padi sebagai variabel respon (y), dan luas lahan (x_1), jumlah petani (x_2), jumlah curah hujan (x_3), penyinaran matahari (x_4), kelembaban (x_5), dan temperatur (x_6) menjadi variabel prediktor.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dibahas di atas, dapat dirumuskan permasalahan sebagai berikut:

- a. bagaimana membuat *interface* web interaktif menggunakan R-Shiny untuk analisis model aditif tergeneralisasi dengan penghalus *spline* yang memiliki kemampuan seperti menggunakan R dengan pendekatan skrip?
- b. bagaimana mengaplikasikan program web interaktif menggunakan R-Shiny untuk menganalisis model aditif tergeneralisasi dengan penghalus *spline* pada data produktivitas padi dan interpretasi hasilnya?

1.3 Batasan Masalah

Adapun batasan masalah dalam penelitian ini adalah model aditif tergeneralisasi yang digunakan termasuk dalam model semiparametrik, dan basis *spline* yang digunakan adalah *cyclic cubic regression spline* (“cc”).

1.4 Tujuan

Berdasarkan rumusan masalah yang dibahas, dapat diketahui tujuan dari penelitian ini, yaitu:

- a. membuat *interface* web interaktif menggunakan R-Shiny untuk analisis model aditif tergeneralisasi dengan penghalus *spline* yang memiliki kemampuan seperti menggunakan R dengan pendekatan skrip tetapi lebih mudah digunakan;
- b. mengaplikasikan program web interaktif menggunakan R-Shiny untuk menganalisis model aditif tergeneralisasi dengan penghalus *spline* pada data produktivitas padi dan menginterpretasi hasilnya.

1.5 Manfaat

Penelitian ini diharapkan dapat membantu dan mempermudah seseorang yang kurang mengerti ilmu statistika maupun program statistika dalam melakukan analisis model aditif tergeneralisasi dengan penghalus *spline* hanya dengan memasukkan data yang akan dianalisis pada program web interaktif secara *online*.

BAB 2. TINJAUAN PUSTAKA

2.1 Model Linier

Perkembangan model linier dimulai dengan perkembangan analisis regresi pada abad ke 19 oleh Pearson, dilanjutkan dengan perkembangan korelasi setelah itu. Teori regresi ini yang menjadi dasar perkembangan teori model linier (Tirta, 2009). Menurut Ratih dalam Handayani (2006), model linier memegang peranan penting dalam suatu percobaan yang melibatkan lebih dari satu variabel yaitu melakukan analisis hubungan antarvariabel. Model linier paling sederhana yang sudah dikenal adalah model regresi sederhana yang hanya melibatkan satu variabel respon dan satu variabel prediktor. Prediktor pada model linier disebut sebagai prediktor linier.

Bentuk umum model linier adalah sebagai berikut:

$$Y = \alpha + \sum_{j=1}^p \beta_j X_j + \varepsilon \quad (2.1)$$

dengan Y adalah variabel respon; X_1, X_2, \dots, X_p adalah variabel prediktor; dan $\alpha, \beta_1, \dots, \beta_p$ merupakan koefisien linier. Nilai koefisien linier dapat diestimasi menggunakan metode kuadrat terkecil. Variabel respon pada model linier diasumsikan berdistribusi normal dan mempunyai fungsi linier.

2.2 Model Linier Tergeneralisasi

Model linier tergeneralisasi (*Generalized Linear Models* atau GLM) merupakan pengembangan model linier ke dalam bentuk distribusi eksponensial. Model linier tergeneralisasi memiliki asumsi model yang lebih fleksibel dari model linier dan mengandung tiga hal berikut:

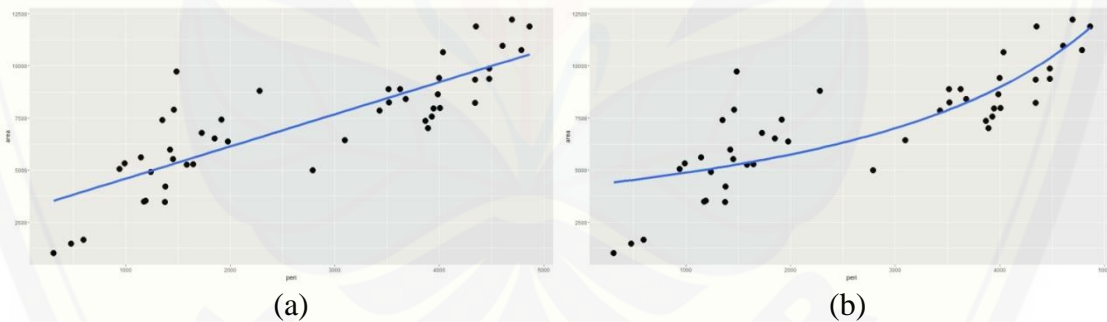
1. komponen tetap yang disebut prediktor linier $\eta_i = \sum_{j=1}^p \beta_j X_{ij}$;
2. respon y_i berdistribusi secara independen dalam keluarga eksponensial;
3. hubungan antara mean dengan prediktor linier ditunjukkan dengan fungsi monoton dan diferensiabel, $g(\cdot)$, yang disebut fungsi hubungan (*link-function*) sedemikian hingga $g(\mu_i) = \eta_i$.

Terdapat fungsi hubungan khusus yang disebut fungsi hubungan kanonik atau natural yang berkaitan erat dengan distribusi y . Misalnya, jika variabel respon berdistribusi normal maka $g(\cdot)$ adalah identitas sehingga $\mu_i = \eta_i$. Dari ketiga hal di atas dapat dikatakan bahwa komponen penting dalam model linier tergeneralisasi ada tiga yaitu:

1. adanya prediktor linier;
2. adanya distribusi keluarga eksponensial;
3. adanya fungsi hubungan

(Tirta, 2009).

Gambar 2.1 berikut menunjukkan perbedaan pencocokan antara model linier dengan model linier tergeneralisasi pada suatu sebaran data. Pada gambar terlihat bahwa pencocokan model menggunakan model linier tergeneralisasi lebih pas memenuhi sebaran data yang ada jika dibandingkan dengan pencocokan model menggunakan model linier. Namun, masih terdapat data yang tidak dapat didekati dengan baik oleh pencocokan model menggunakan model linier tergeneralisasi.



(a) Menggunakan model linier; (b) Menggunakan model linier tergeneralisasi

Gambar 2.1 Contoh pencocokan model

2.3 Model Aditif

Menurut Hastie dan Tibshirani (1990), model aditif (*Additive Models*) merupakan hasil generalisasi atau pengembangan dari model linier. Sama halnya dengan model linier, variabel respon pada model aditif juga diasumsikan berdistribusi normal. Perbedaan kedua model tersebut terletak pada fungsi yang dimiliki masing-

masing model. Jika pada model linier mengharuskan fungsinya linier, maka pada model aditif tidak harus linier. Model linier digeneralisir menjadi model aditif dengan mengganti prediktor linier $\eta(X) = \sum_{j=1}^p \beta_j X_j$, menjadi prediktor aditif $\eta(X) = \sum_{j=1}^p f_j(X_j)$, sehingga model aditif dapat dituliskan sebagai berikut:

$$Y = \alpha + \sum_{j=1}^p f_j(X_j) + \varepsilon \quad (2.2)$$

dengan Y adalah variabel respon; X_1, X_2, \dots, X_p adalah variabel prediktor; f_1, f_2, \dots, f_p adalah fungsi dari variabel prediktor; dan ε adalah galat.

Model aditif disebut sebagai model nonparametrik jika semua fungsi f_j pada prediktor aditif $\eta(X)$, merupakan fungsi mulus s_j . Sehingga model aditif juga dapat dituliskan ke dalam bentuk model nonparametrik sebagai berikut:

$$Y = s_0 + \sum_{j=1}^p s_j(X_j) + \varepsilon \quad (2.3)$$

Fungsi mulus s_j memenuhi kondisi standar yaitu $E[s_j(X_j)] = 0$. Model aditif juga dapat dituliskan ke dalam bentuk model semiparametrik $Y = \alpha + \sum \beta_j X_j + \gamma(t) + \varepsilon$, dengan suku pertama $\alpha + \sum \beta_j X_j$, sebagai bagian parametriknya dan suku kedua $\gamma(t)$, sebagai bagian nonparametriknya.

2.4 Model Aditif Tergeneralisasi

Model aditif tergeneralisasi (*Generalized Additive Models*) mengembangkan model aditif ke dalam bentuk distribusi keluarga eksponensial. Model aditif tergeneralisasi juga dapat disebut sebagai model pengembangan dari model linier tergeneralisasi, karena model ini mengembangkan model linier tergeneralisasi dengan cara yang sama dengan model aditif mengembangkan model linier, yaitu mengganti prediktor linier $\sum_{j=1}^p \beta_j X_j$ dengan prediktor aditif $\sum_{j=1}^p f_j(X_j)$. Hal tersebut membuat model aditif tergeneralisasi menjadi lebih fleksibel daripada model linier tergeneralisasi maupun model aditif.

Seperti pada model linier tergeneralisasi, model aditif tergeneralisasi juga memuat komponen acak, komponen tetap yaitu prediktor aditif, dan fungsi *link* yang

menghubungkan dua komponen tersebut. Variabel respon Y sebagai komponen acak diasumsikan memiliki fungsi kepadatan keluarga eksponensial sebagai berikut:

$$f_Y(y; \theta; \phi) = \exp \left\{ \frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi) \right\} \quad (2.4)$$

dengan θ adalah parameter natural dan ϕ adalah parameter skala. Hubungan antara mean μ dari variabel respon dan komponen aditif $\eta(X) = s_0 + \sum_{j=1}^p s_j(X_j)$ didefinisikan oleh fungsi *link* $g(\mu) = \eta$ dan secara umum fungsi *link* yang paling umum digunakan adalah *link* kanonik dengan $\eta = \theta$ (Hastie dan Tibshirani, 1990). Model aditif tergeneralisasi dapat diformulasikan sebagai berikut:

$$g(\mu) = s_0 + \sum_{j=1}^p s_j(X_j) + \varepsilon \quad (2.5)$$

dengan variabel responnya diasumsikan berdistribusi keluarga eksponensial, dan $s(\cdot)$ merupakan fungsi mulus.

Model aditif tergeneralisasi dan model linier tergeneralisasi dapat diterapkan dalam situasi yang mirip, tetapi kedua model tersebut memberikan tujuan analitik yang berbeda. Model linier tergeneralisasi menekankan estimasi dan inferensi untuk parameter model, sementara model aditif tergeneralisasi fokus pada eksplorasi data nonparametrik (Statsoft, 2016).

2.5 Penghalus *Spline*

Pada model aditif tergeneralisasi, fungsi atau komponen nonparametrik diestimasi oleh fungsi penghalus yakni penghalus diagram pencar (*scatterplot smoother*). Hastie dan Tibshirani (1990), menjelaskan berbagai macam penghalus diagram pencar, salah satunya adalah penghalus *spline*. Misalkan diberikan data $\{(x_i, y_i), i = 1, \dots, n\}$ dan hubungan antara x_i dan y_i diasumsikan mengikuti model regresi nonparametrik (2.6) berikut

$$y_i = f(x_i) + \varepsilon_i, x_i \in [a, b] \quad (2.6)$$

dengan $f(x_i)$ adalah fungsi yang tidak diketahui bentuknya dan hanya diasumsikan mulus. Menurut Herawati (2011), penghalus *spline* akan mengestimasi fungsi f

sebagai solusi dari masalah optimasi yaitu dengan mencari $\hat{f} \in L_2[a, b]$ yang meminimumkan persamaan *penalized least squares* (2.7) berikut

$$S(f) = \sum_{i=1}^n (y_i - f(x_i))^2 + \lambda \int_a^b (f''(x))^2 dx \quad (2.7)$$

dengan $L_2[a, b]$ menyatakan himpunan fungsi-fungsi kuadrat terintegral pada interval $[a, b]$; λ adalah nilai konstan (*fixed constant*); dan $a \leq x_1 \leq \dots \leq x_n \leq b$. Suku pertama pada persamaan (2.7) digunakan untuk mengukur kerapatan data, yang tidak lain merupakan jumlah kuadrat sisaan, sedangkan suku kedua merupakan penalti kekasaran (*roughness penalty*) yang memberikan ukuran kemulusan atau kekasaran kurva dalam memetakan data. Suku kedua tersebut akan mempenalti kurva dari fungsi f . Fungsi linier akan memiliki nilai $\int (f''(x))^2 dx = 0$, sedangkan fungsi nonlinier memiliki nilai $\int (f''(x))^2 dx > 0$. Parameter λ merupakan parameter penghalus (*smoothing parameter*) yang bernilai positif ($0 < \lambda$). Jika nilai $\lambda \rightarrow +\infty$, maka penalti mendominasi persamaan (2.6) dan penduga *spline* menjadi konstan, serta akan memberikan kurva yang mulus. Sedangkan jika nilai $\lambda \rightarrow 0$, maka penalti tidak lagi ada dalam persamaan (2.7) dan penduga *spline* menginterpolasi data, serta akan memberikan kurva yang kasar. Sehingga, parameter penghalus λ memainkan peran sentral dalam mengendalikan keseimbangan (*trade off*) antara ketepatan model (*goodness of fit*) dan kemulusan kurva (Herawati, 2011).

Salah satu penghalus *spline* adalah *spline* kubik dengan titik perubahan yang terjadi di dalam suatu kurva yang disebut titik knot. Fungsi *spline* berorde m dengan titik-titik knot K_1, K_2, \dots, K_N secara umum dapat disajikan dalam bentuk (Herawati, 2011):

$$f(x) = \beta_0 + \sum_{j=1}^p \left[\sum_{r=1}^{m-1} \beta_{j,r} X_j^r + \sum_{k=1}^N \beta_{j,(m-1),k} (X_j - K_{jk})_+^{m-1} \right] \quad (2.8)$$

dengan

$$(X_j - K_{jk})_+^{m-1} = \begin{cases} (X_j - K_{jk})^{m-1} & , X_j \geq K_{jk} \\ 0 & , X_j < K_{jk} \end{cases}$$

Jika pada persamaan (2.8) diambil nilai $m = 4$, maka diperoleh *spline* kubik. Menurut Green dan Silverman dalam Herawati (2011), suatu fungsi $f(x)$ disebut

penghalus *spline* kubik jika memenuhi dua kondisi, yaitu kondisi pertama, fungsi $f(x)$ pada setiap interval (a, x_1) , (x_1, x_2) , (x_2, x_3) , ..., (x_n, b) adalah polinomial kubik, dan kondisi kedua yaitu fungsi $f(x)$ kontinu pada turunan pertama dan kedua pada setiap $x_i \in [a, b]$ dimana x_i merupakan titik knot.

2.5.1 Regresi *Spline* Kubik

Misalkan terdapat sebuah model yang mengandung satu fungsi mulus seperti pada model (2.6) berikut

$$y_i = f(x_i) + \varepsilon_i$$

dengan $i = 1, \dots, n$ dan n adalah banyaknya pengamatan. Estimasi fungsi mulus f dengan regresi *spline* kubik dapat dinyatakan ke dalam bentuk

$$f(x) = \sum_{j=1}^N h_j(x) \beta_j \quad (2.9)$$

dimana N adalah banyaknya knot; $h_j(x)$ merupakan fungsi basis ke- j untuk regresi *spline* kubik; dan β_j adalah parameter yang tidak diketahui (Wood, 2006). Menurut Wood (2006), untuk sebarang fungsi f yang mempunyai sebuah ekspansi basis seperti pada (2.9), memungkinkan untuk menulis penalti kekasaran (*roughness penalty*) seperti berikut

$$\int_{x_1}^{x_N} (f''(x))^2 dx = \boldsymbol{\beta}^T \mathbf{S} \boldsymbol{\beta} \quad (2.10)$$

dengan x_1, \dots, x_N adalah titik knot; $\boldsymbol{\beta}^T = (\beta_1, \dots, \beta_N)$; dan \mathbf{S} adalah matriks koefisien yang diketahui. Matriks penalti untuk basis regresi *spline* kubik ini adalah

$$\mathbf{S} = \mathbf{D}^T \mathbf{B}^{-1} \mathbf{D}$$

dengan \mathbf{D} adalah matriks *upper triangular* dan \mathbf{B} adalah matriks *symmetric tridiagonal* yang didefinisikan pada Tabel 2.1.

Tabel 2.1 Definisi matriks untuk regresi *spline* kubik

Elemen matriks tak nol – <i>spline</i> tak siklik		
$D_{j,j} = 1/h_j$	$D_{j,j+1} = -1/h_j - 1/h_{j+1}$	$D_{j,j+2} = 1/h_{j+1}$
$B_{j,j} = (h_j + h_{j+1})/3$		$j = 1 \dots N - 2$
$B_{j,j+1} = h_{j+1}/6$	$B_{j+1,j} = h_{j+1}/6$	$j = 1 \dots N - 3$

dengan $h_j = x_{j+1} - x_j$.
 Sumber: Wood (2006).

Fungsi regresi *spline* kubik pada (2.9) dapat dituliskan ke dalam bentuk matriks yaitu

$$\mathbf{f} = \mathbf{H}\boldsymbol{\beta} \tag{2.11}$$

dengan $\mathbf{f} = (f(x_1), \dots, f(x_n))^T$; $\boldsymbol{\beta} = (\beta_1, \dots, \beta_N)^T$; dan $H_{i,j} = h_j(x_i)$. Nilai estimasi parameter β pada fungsi mulus f diperoleh dengan meminimumkan persamaan (2.7) yaitu

$$S(f) = \sum_{i=1}^n (y_i - f(x_i))^2 + \lambda \int_a^b (f''(x))^2 dx$$

Persamaan (2.7) dapat dituliskan dalam bentuk matriks seperti berikut

$$S(f) = (\mathbf{y} - \mathbf{H}\boldsymbol{\beta})^T (\mathbf{y} - \mathbf{H}\boldsymbol{\beta}) + \lambda \boldsymbol{\beta}^T \mathbf{S}\boldsymbol{\beta} \tag{2.12}$$

Syarat perlu agar persamaan di atas minimum adalah $\frac{\delta S(f)}{\delta \boldsymbol{\beta}} = 0$, sehingga diperoleh

nilai estimasi untuk parameter β yaitu

$$\boldsymbol{\beta} = (\mathbf{H}^T \mathbf{H} + \lambda \mathbf{S})^{-1} \mathbf{H}^T \mathbf{y} \tag{2.13}$$

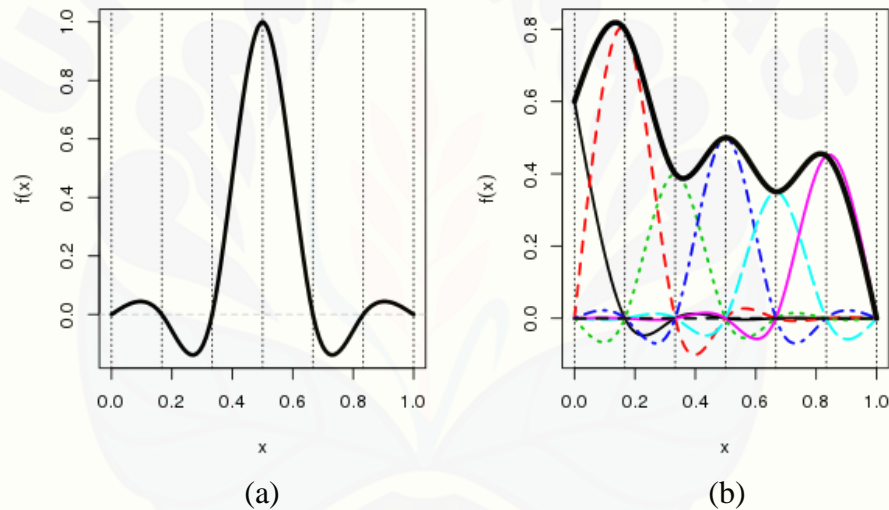
dengan mensubstitusikan persamaan (2.13) ke (2.11), maka didapatkan estimasi fungsi mulus f dengan penghalus regresi *spline* kubik yaitu

$$\mathbf{f} = \mathbf{H}(\mathbf{H}^T \mathbf{H} + \lambda \mathbf{S})^{-1} \mathbf{H}^T \mathbf{y} \tag{2.14}$$

dengan $\mathbf{S} = \mathbf{D}^T \mathbf{B}^{-1} \mathbf{D}$.

Gambar 2.2 merupakan ilustrasi basis regresi *spline* kubik dalam membangun estimasi fungsi mulus. Pada Gambar 2.2, gambar (a) menunjukkan satu fungsi basis yang memusat pada sebuah knot. Garis vertikal putus-putus menunjukkan lokasi

knot. Semakin banyak knot yang dimiliki, maka semakin kompleks fungsi *spline* yang didapatkan. Fungsi basis *spline* kubik untuk sebuah knot yang diberikan mengambil nilai 1 pada lokasi knotnya, dan nilai 0 pada semua knot lainnya. Gambar (b) menunjukkan bagaimana fungsi-fungsi basis dikombinasikan untuk menggambarkan estimasi fungsi mulus. Berbagai kurva dengan warna yang berbeda menunjukkan fungsi basis $h_j(x)$ dari regresi *spline* kubik, masing-masing dikalikan dengan koefisien β_j yang berkorespondensi, kemudian dijumlahkan untuk mendapatkan estimasi fungsi mulus yang diilustrasikan dengan kurva kontinu tebal berwarna hitam.



Gambar 2.2 Ilustrasi basis regresi *spline* kubik (Sumber: Wood, 2006)

2.5.2 Regresi *Spline* Kubik Siklik

Regresi *spline* kubik siklik adalah spesifikasi dari regresi *spline* kubik yang nilai fungsi dan turunannya hingga order kedua di dua titik ujungnya adalah sama. Seperti regresi *spline* kubik yang telah dijelaskan sebelumnya, estimasi fungsi mulus f pada model (2.6) dengan regresi *spline* kubik siklik juga dapat dinyatakan ke dalam bentuk (2.15) berikut

$$f(x) = \sum_{j=1}^{N-1} g_j(x)\beta_j \quad (2.15)$$

dengan N adalah banyaknya knot; $g_j(x)$ merupakan fungsi basis ke- j untuk regresi *spline* kubik siklik; dan β_j adalah parameter yang tidak diketahui (Wood, 2006). Berbeda dengan regresi *spline* kubik sebelumnya, pada regresi *spline* kubik siklik $\beta_1 = \beta_N$ dan $g_1(x) = g_N(x)$, karenanya j pada persamaan (2.15) bergerak dari 1 hingga $N - 1$. Fungsi f pada (2.15) mempunyai sebuah ekspansi basis, sehingga memungkinkan untuk menulis penalti kekasaran (*roughness penalty*) seperti pada (2.10) berikut

$$\int_{x_1}^{x_N} (f''(x))^2 dx = \boldsymbol{\beta}^T \mathbf{S} \boldsymbol{\beta}$$

dalam hal ini $\boldsymbol{\beta}^T = (\beta_1, \dots, \beta_{N-1})$; dan \mathbf{S} adalah matriks koefisien yang diketahui. Matriks penalti untuk basis regresi *spline* kubik siklik ini adalah

$$\mathbf{S} = \mathbf{M}^T \mathbf{N}^{-1} \mathbf{M}$$

dengan \mathbf{M} dan \mathbf{N} adalah matriks yang didefinisikan pada Tabel 2.2.

Tabel 2.2 Definisi matriks untuk regresi *spline* kubik siklik

Elemen matriks tak nol – <i>spline</i> siklik		
$N_{j-1,j} = N_{j,j-1} = h_{j-1}/6$	$N_{j,j} = (h_{j-1} + h_j)/3$	
$M_{j-1,j} = M_{j,j-1} = 1/h_{j+1}$	$M_{j,j} = -1/h_{j-1} - 1/h_j$	$j = 2 \dots N - 1$
$N_{1,1} = (h_{k-1} + h_1)/3$	$N_{1,k-1} = h_{k-1}/6$	$N_{k-1,1} = h_{k-1}/6$
$M_{1,1} = -1/h_1 - 1/h_{k-1}$	$M_{1,k-1} = 1/h_{k-1}$	$M_{k-1,1} = 1/h_{k-1}$

Sumber: Wood (2006).

Seperti pada regresi *spline* kubik sebelumnya, fungsi regresi *spline* kubik siklik pada (2.15) juga dapat dituliskan ke dalam bentuk matriks, yaitu

$$\mathbf{f} = \mathbf{G} \boldsymbol{\beta} \tag{2.16}$$

dengan $\mathbf{f} = (f(x_1), \dots, f(x_n))^T$; $\boldsymbol{\beta} = (\beta_1, \dots, \beta_{N-1})^T$; dan $G_{i,j} = g_j(x_i)$. Dengan cara yang sama dapat diperoleh nilai estimasi untuk parameter β yaitu

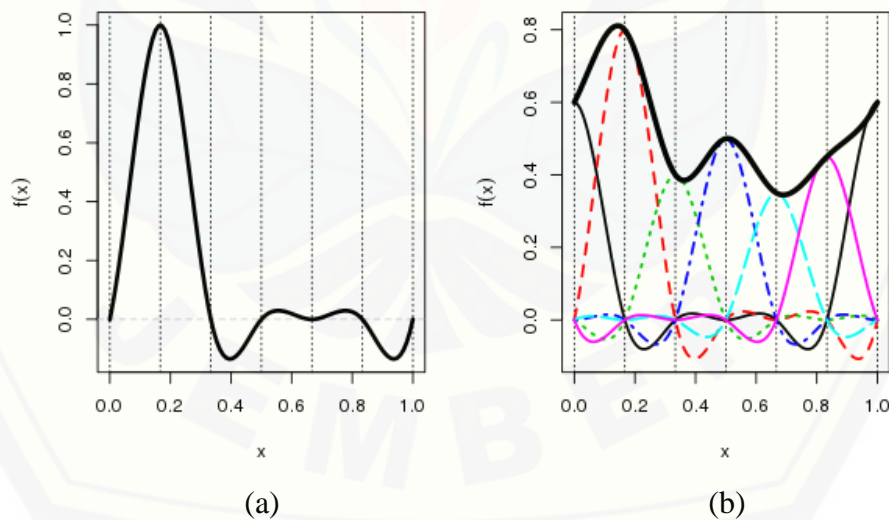
$$\boldsymbol{\beta} = (\mathbf{G}^T \mathbf{G} + \lambda \mathbf{S})^{-1} \mathbf{G}^T \mathbf{y} \tag{2.17}$$

dengan mensubstitusikan persamaan (2.17) ke (2.16), maka didapatkan estimasi fungsi mulus f dengan penghalus regresi *spline* kubik siklik yaitu

$$\mathbf{f} = \mathbf{G}(\mathbf{G}^T \mathbf{G} + \lambda \mathbf{S})^{-1} \mathbf{G}^T \mathbf{y} \quad (2.18)$$

dengan $\mathbf{S} = \mathbf{M}^T \mathbf{N}^{-1} \mathbf{M}$.

Gambar 2.3 merupakan ilustrasi basis regresi *spline* kubik siklik dalam membangun estimasi fungsi mulus. Pada Gambar 2.3, gambar (a) menunjukkan satu fungsi basis yang memusat pada sebuah knot. Gambar (b) menunjukkan bagaimana fungsi-fungsi basis dikombinasikan untuk menggambarkan estimasi fungsi mulus. Berbagai kurva dengan warna yang berbeda menunjukkan fungsi basis $g_j(x)$ dari regresi *spline* kubik, masing-masing dikalikan dengan koefisien β_j yang berkorespondensi, kemudian dijumlahkan untuk mendapatkan estimasi fungsi mulus yang diilustrasikan dengan kurva kontinu tebal berwarna hitam. Terlihat pada gambar (b), estimasi fungsi mulus memiliki nilai fungsi yang sama di kedua titik ujungnya.



Gambar 2.3 Ilustrasi basis regresi *spline* kubik siklik (Sumber: Wood, 2006)

Berdasarkan uraian dan ilustrasi di atas, penghalus regresi *spline* kubik siklik dapat digunakan untuk mengestimasi fungsi regresi *smooth* yang periodik, seperti pola suhu harian atau tahunan.

2.6 P-IRLS

Menurut Wood (2006), estimasi model aditif tergeneralisasi menggunakan maksimum *likelihood* terpenalti (*penalized likelihood maximization*) yang dapat dicapai dengan menggunakan algoritma *Penalized Iteratively Re-weighted Least Square* (P-IRLS). Sketsa dasar dari P-IRLS dalam mengestimasi model aditif tergeneralisasi ditunjukkan sebagai berikut:

1. Diberikan $\boldsymbol{\beta}^{[k]}$ yang merupakan vektor parameter yang diestimasi pada iterasi ke- k , dan $\boldsymbol{\eta}^{[k]}$ serta $\boldsymbol{\mu}^{[k]}$ yang merupakan vektor dengan elemen $\eta_i^{[k]} = \mathbf{X}_i \boldsymbol{\beta}^{[k]}$ dan $\mu_i^{[k]} = g^{-1}(\eta_i^{[k]})$, dimana \mathbf{X}_i adalah baris ke- i dari matriks model \mathbf{X} dan $g^{-1}(\cdot)$ adalah fungsi invers dari *link*. Menghitung pseudodata $\mathbf{z}^{[k]}$ dengan elemen

$$z_i^{[k]} = g'(\mu_i^{[k]})(y_i - \mu_i^{[k]}) + \eta_i^{[k]} \quad (2.19)$$

dan bobot $\mathbf{W}^{[k]}$ dengan elemen

$$W_{ii}^{[k]} = \frac{1}{v(\mu_i^{[k]})g'(\mu_i^{[k]})^2} \quad (2.20)$$

dimana Z adalah variabel dependen yang digunakan pada model aditif tergeneralisasi, merupakan variabel dependen yang telah disesuaikan dengan fungsi *link* yang bersesuaian dengan distribusi dari variabel respon Y ; \mathbf{W} adalah matriks diagonal; dan $var(Y_i) = V(\mu_i^{[k]})\phi$, dengan ϕ adalah parameter skala yang berubah-ubah.

2. Mendapatkan nilai $\boldsymbol{\beta}^{[k+1]}$ dengan meminimumkan

$$\|\sqrt{\mathbf{W}^{[k]}}(\mathbf{z}^{[k]} - \mathbf{X}\boldsymbol{\beta})\|^2 + \boldsymbol{\beta}^T \mathbf{S}\boldsymbol{\beta} \quad (2.21)$$

terhadap $\boldsymbol{\beta}$. Kenaikan iterasi $k = k + 1$, iterasi dilakukan hingga konvergen.

2.7 Pemilihan Parameter Penghalus

Menurut Budiantara (dalam Tripena, 2011), bentuk estimator *spline* sangat dipengaruhi oleh nilai parameter penghalus λ . Pemilihan λ optimal dalam regresi *spline* pada hakikatnya merupakan pemilihan lokasi titik-titik knot. Mengingat tujuan

dari pendekatan regresi nonparametrik, yaitu ingin didapatkan kurva mulus yang mempunyai λ optimal menggunakan data amatan sebanyak n , maka diperlukan ukuran kinerja atas estimator yang dapat diterima secara universal. Eubank (1999) menyebutkan bahwa, ukuran kinerja atas estimator tersebut adalah *Generalized Cross-Validation* (GCV). GCV merupakan modifikasi dari *Cross-Validation* (CV) atau validasi silang, persamaan validasi silang didefinisikan sebagai berikut:

$$CV(\lambda) = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{f}_\lambda^{-1}(x_i))^2 \quad (2.22)$$

dengan $\hat{f}_\lambda^{-1}(x_i)$ menyatakan kecocokan pada saat x_i dan $CV(\lambda)$ disebut sebagai prediksi jumlah kuadrat (Hastie dan Tibshirani, 1990). Semua penghalusan yang diestimasi dapat dirumuskan sebagai kombinasi linier dari:

$$\hat{\mathbf{f}}_\lambda = \mathbf{S}_\lambda \mathbf{y}$$

dengan \mathbf{S}_λ adalah matriks penghalus yang berukuran $n \times n$. Misalkan $S_{ii}(\lambda)$ adalah elemen diagonal dari \mathbf{S}_λ maka persamaan validasi silang dapat dituliskan sebagai:

$$CV(\lambda) = \frac{1}{n} \sum_{i=1}^n \left(\frac{y_i - \hat{f}_\lambda(x_i)}{1 - S_{ii}(\lambda)} \right)^2 \quad (2.23)$$

Dengan mengganti $S_{ii}(\lambda)$ dengan $tr(\mathbf{S}_\lambda)/n$ pada persamaan (2.23) akan didapatkan validasi silang tergeneralisasi atau *Generalized Cross Validation* (GCV) yang diformulasikan sebagai berikut:

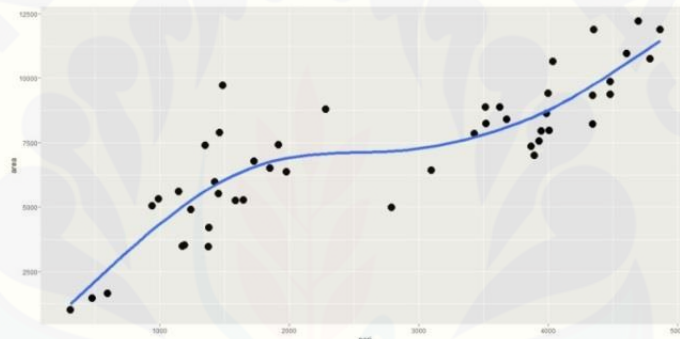
$$GCV(\lambda) = \frac{1}{n} \sum_{i=1}^n \left(\frac{y_i - \hat{f}_\lambda(x_i)}{1 - tr(\mathbf{S}_\lambda)/n} \right)^2 \quad (2.24)$$

Kriteria GCV diharapkan memiliki nilai yang minimum, sehingga model regresi *spline* dapat dikatakan memiliki nilai λ yang optimal. Selain itu, nilai statistik AIC (*Akaike's Information Criterion*) juga diperlukan untuk memilih ketepatan atau kecocokan model secara keseluruhan. Besarnya AIC dapat dilihat pada persamaan berikut

$$AIC = -2l(\hat{\boldsymbol{\theta}}) + 2q \quad (2.25)$$

dengan $l(\hat{\theta})$ adalah nilai *likelihood* dari model dan q adalah banyaknya parameter dalam model. Nilai AIC yang lebih kecil menunjukkan model yang lebih tepat atau cocok (Tirta, 2009).

Gambar 2.4 berikut menunjukkan pencocokan model menggunakan model aditif tergeneralisasi dengan penghalus *spline*. Pada gambar terlihat bahwa pencocokan model menggunakan model aditif tergeneralisasi dengan penghalus *spline* lebih pas dalam memenuhi sebaran data yang ada jika dibandingkan dengan pencocokan model menggunakan model linier maupun model linier tergeneralisasi (lihat Gambar 2.1).



Gambar 2.4: Pencocokan model menggunakan model aditif tergeneralisasi dengan penghalus *spline*

2.8 R-Shiny

R-Shiny merupakan bagian dari program R yang dapat membuat program secara *online*. Secara umum, komponen program R-Shiny dibedakan menjadi dua kelompok besar yaitu (Tirta, 2014):

1. *User Interface*. Bagian ini bermanfaat untuk:
 - a. *panel control*, yaitu panel untuk mengontrol input berupa data, variabel, model, tergantung kompleksitas modul. Tampilan kontrol dapat berupa *slider*, *radio button*, *check-box*, dan lain-lain
 - b. pemasukan permintaan nilai *input* (data dengan berbagai jenis variabel yang diperlukan, pemilihan model, jenis, dan kriteria uji statistika)
 - c. penyajian *output* terkait hasil analisis/uji. Hasil *output* dapat berupa

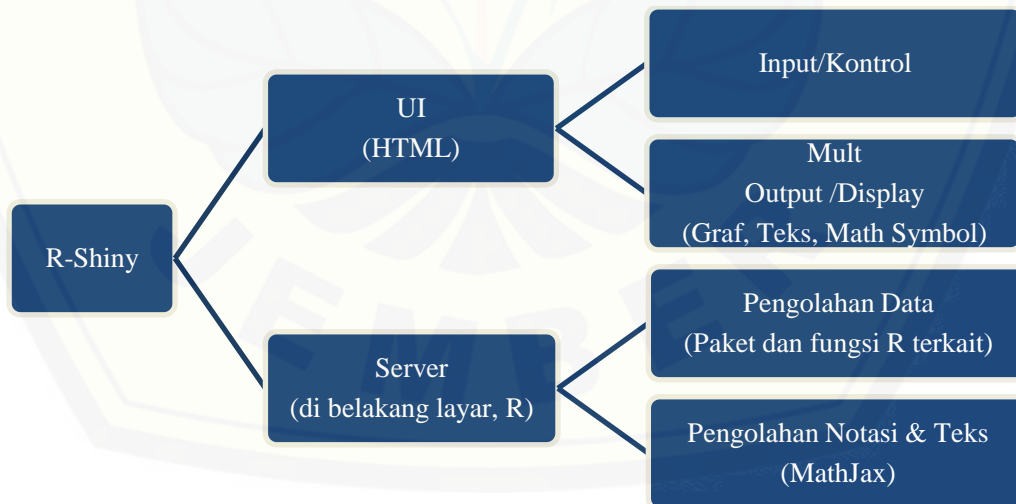
- i. grafik (histogram, diagram pencar, dan lain-lain);
- ii. bentuk angka/teks bisa berbentuk asli (verbatim) maupun dalam bentuk tabel;
- iii. teks khusus dengan notasi matematika dengan format LaTeX.

Untuk mengakomodasi berbagai jenis luaran tadi, R-Shiny juga menyediakan berbagai format *output* seperti *plotOutput*, *textOutput*, *verbatimTextOutput*, *tableOutput*, dan lain-lainnya.

Bagian *user interface* ini dapat disajikan pada file khusus *ui.r*, dapat juga disajikan penuh melalui file HTML, misalnya *index.html*.

2. *Server*. Bagian ini merupakan otak dari program yang bertugas melakukan simulasi berbagai analisis data sesuai pilihan pengguna dan selanjutnya mengirim hasilnya ke bagian *output*. Bagian ini didukung oleh berbagai prosedur dan analisis data yang pada umumnya telah tersedia pada berbagai paket R. Bagian ini disimpan dalam file yang diberi nama *server.r*.

Struktur umum komponen pemrograman dengan R-Shiny dapat dilihat pada Gambar 2.5.



Gambar 2.5 Struktur komponen R-Shiny

Secara umum, penyajian file ui dan file server ada dua macam yaitu (Tirta, 2014):

1. membuat file server.r dan ui.r secara terpisah lalu diletakkan pada satu direktori yang diberi nama sesuai dengan topiknya. Selanjutnya yang dipanggil adalah nama direktorinya.

Direktori: shiny1

File:

```
-server.r  
/subdir www/  
-index.html
```

Cara memanggil: runApp()

2. menggabungkan file server dan ui dalam satu fungsi. Selanjutnya pemanggilan dilakukan dengan memanggil nama fungsinya.

```
Fungsi.shiny1<-function() {  
  ui=,  
  server=  
}
```

Cara memanggil: fungsi.shiny1()

3. untuk menampilkan rumus-rumus matematika di dalam tampilan R-shiny, dapat dilakukan dengan memanfaatkan MathJax yang memiliki sintaks mirip LaTeX.

```
WithMathJax(  
  helpText("Text biasa $$simbol matematika $$"))
```


BAB 3. METODE PENELITIAN

Metode penelitian merupakan kerangka acuan yang akan digunakan dalam menganalisis suatu data serta menerangkan tentang langkah-langkah kerja yang harus dilakukan agar menghasilkan hasil analisis yang diinginkan.

3.1 Data

Data yang digunakan pada penelitian ini adalah data sekunder yang berasal dari Badan Pusat Statistik (BPS) yang diambil dari laman <http://www.bps.go.id/> berupa data iklim dan analisis hasil sensus pertanian di Indonesia tahun 2013. Unit observasi yang digunakan dalam penelitian ini adalah provinsi di Indonesia. Adapun variabel yang digunakan dalam penelitian ini yaitu:

- a. data produktivitas padi sawah yang menjadi variabel respon (y),
- b. data luas lahan (x_1), jumlah petani (x_2), jumlah curah hujan (x_3), penyinaran matahari (x_4), kelembaban (x_5), dan temperatur (x_6) menjadi variabel prediktor.

3.2 Langkah-langkah Penelitian

Langkah-langkah penelitian analisis model aditif tergeneralisasi dengan penghalus *spline* berbasis web interaktif menggunakan R-Shiny dibagi menjadi dua tahap, yaitu tahap penyusunan web interaktif dan tahap aplikasi program web interaktif untuk analisis data.

3.2.1 Tahap Penyusunan Web Interaktif

Tahap pertama yaitu penyusunan web interaktif. Langkah-langkah penyusunan web interaktif adalah sebagai berikut:

1. Mengidentifikasi teori, *input* dan *output* model aditif tergeneralisasi dengan penghalus *spline*. Jenis penghalus *spline* yang akan digunakan pada penelitian ini adalah penghalus regresi *spline* kubik siklik, sehingga teori yang akan dibahas

adalah model aditif tergeneralisasi dengan penghalus regresi *spline* kubik siklik. *Input* yang akan dientrikan berupa *input* data dan *input* model berupa variabel x , variabel y , distribusi beserta fungsi *link*, basis *spline*, dan dimensi basis (N). Sedangkan *output* yang akan dihasilkan berupa *summary*, *scatterplot*, dan histogram distribusi untuk eksplorasi data, *summary* dan *plotting* untuk estimasi parameter, serta nilai AIC untuk penentuan model terbaik. Paket dalam R yang digunakan pada penelitian ini yaitu *mgcv*. Adapun struktur fungsi model aditif tergeneralisasi dalam paket *mgcv* sebagai berikut:

```
gam(formula, family=..., data=..., method="GCV.Cp", scale=0,  
     knots=NULL, sp=NULL, min.sp=NULL, gamma=1, ...)
```

uraian mengenai perintah-perintah pada fungsi di atas adalah sebagai berikut:

formula :suatu rumusan dari model regresi dalam bentuk $\text{response} \sim \text{predictors}$, di dalamnya terdapat fungsi untuk penghalus nonparametrik yang diindikasikan oleh s untuk penghalus *spline*;

family :deskripsi dari data *error* dan fungsi *link* yang digunakan dalam model;

data :suatu data frame yang memuat variabel-variabel dalam model, variabel-variabel tersebut diambil dari `environment(formula)`;

method :metode estimasi parameter penghalus yang digunakan. Jika bernilai "GCV.Cp", maka menggunakan GCV untuk parameter skala yang tidak diketahui dan Mallows' Cp/UBRE/AIC untuk parameter skala yang diketahui;

scale :mengontrol apakah menggunakan UBRE atau GCV untuk estimasi penghalusan. Jika bernilai positif maka parameter skala diketahui sehingga digunakan UBRE. Jika bernilai negatif maka parameter skala tidak diketahui sehingga digunakan GCV. Sedangkan jika

bernilai 0 maka digunakan UBRE untuk poisson dan binomial, dan GCV untuk lainnya;

- `knots` :daftar nilai knot yang digunakan untuk membangun basis. Jika bernilai NULL maka nilai knot akan dihasilkan secara otomatis;
- `sp` :sebuah vektor parameter penghalus yang disediakan. Ketika tidak bernilai NULL, elemen yang bernilai negatif mengindikasikan bahwa parameter harus diestimasi;
- `min.sp` :nilai minimum dari parameter penghalus yang disediakan;
- `gamma` :terkadang sangat berguna untuk meningkatkan derajat bebas model pada nilai GCV atau UBRE/AIC dengan menggunakan sebuah pengali konstan;
- `...` :argumen atau perintah lainnya.

Sedangkan struktur fungsi untuk penghalus *spline* dalam paket `mgcv` sebagai berikut:

```
s(..., k=..., bs=..., fx=FALSE, m=NA, by=NA, xt=NULL,
  id=NULL, sp=NULL)
```

uraian mengenai perintah-perintah pada fungsi di atas adalah sebagai berikut:

- `...` :daftar variabel prediktor yang diberi *smoothing*;
- `k` :dimensi basis yang merupakan jumlah titik knot ditambah 2;
- `bs` :mengindikasikan basis penghalus yang digunakan. “cr” untuk *cubic regression spline*, “cc” untuk *cyclic cubic regression spline*, “tp” untuk *thin plate regression spline*, dan lain-lain;
- `fx` :mengindikasikan apakah variabel prediktor diperlakukan sebagai regresi *spline* murni (dengan derajat bebas tetap) (TRUE) atau diperlakukan sebagai regresi *spline* terpenalti (FALSE);
- `m` :orde penalti untuk penghalus *spline*. Jika bernilai NA maka inisialisasi orde dilakukan secara otomatis;
- `by` :nama dari sembarang variabel (NA untuk tidak);

- `xt` :informasi tambahan yang digunakan untuk menyusun sebuah basis khusus;
- `id` :sebuah label atau nomor yang mengidentifikasi penghalus, dan menghubungkannya dengan penghalus lainnya. NULL untuk tidak ada hubungan (*linkage*);
- `sp` :*array* parameter penghalus untuk variabel prediktor (negatif untuk estimasi otomatis).

Adapun fungsi dalam paket `mgcv` yang digunakan untuk memberikan visualisasi atau *plotting* model seperti berikut:

```
plot(x, residuals=TRUE, se=TRUE, pages=1, pers=TRUE,
all.terms=TRUE, seWithMean=TRUE, ... )
```

dengan keterangan:

- `x` :objek pencocokan model;
- `residuals` :jika TRUE maka residu atau sisaan ditambahkan dalam *plot*, jika FALSE, maka residu tidak ditambahkan/ditampilkan;
- `se` :jika TRUE maka garis yang menunjukkan *standard errors* dari estimasi penghalus akan ditampilkan, jika FALSE maka tidak ditampilkan;
- `pages` :jika bernilai 1, maka *plot* akan ditampilkan dalam satu halaman. Jika bernilai 0, maka *plot* akan ditampilkan secara bergantian;
- `pers` :bernilai TRUE jika ingin menggambar *plot* perspektif (3D) untuk model dengan dua penghalusan yang digabung menjadi satu penghalus;
- `all.terms` :jika TRUE, maka komponen model parametrik juga akan digambarkan dalam *plot*;

`seWithMean` :jika TRUE, maka komponen penghalus akan ditampilkan dengan selang kepercayaan yang mencakup ketidakpastian tentang mean secara keseluruhan;
... :argumen atau perintah lainnya.

2. Menyusun dokumen teori model aditif tergeneralisasi dengan penghalus *spline* ke dalam format LaTeX berupa `index.html`, merupakan file HTML (tutorial) yang digunakan untuk pemasukan nilai *input* dan penyajian *output* sehingga dapat ditampilkan secara *online*.
3. Menyusun skrip R-Shiny yang berupa `server.r`. `Server.r` merupakan otak dari program sehingga pada bagian ini dilakukan pengaplikasian program analisis model aditif tergeneralisasi dengan penghalus *spline* pada R ke dalam program R-Shiny. Dalam penelitian ini, untuk menganalisis data model aditif tergeneralisasi dengan penghalus *spline* menggunakan *software* R i386 3.1.2.
4. Mengintegrasikan dan menguji coba dokumen teori dalam `index.html` dan `server.r` yang ada pada langkah 1 dan 2 untuk membuat program *online*.

3.2.2 Tahap Aplikasi Program Web Interaktif

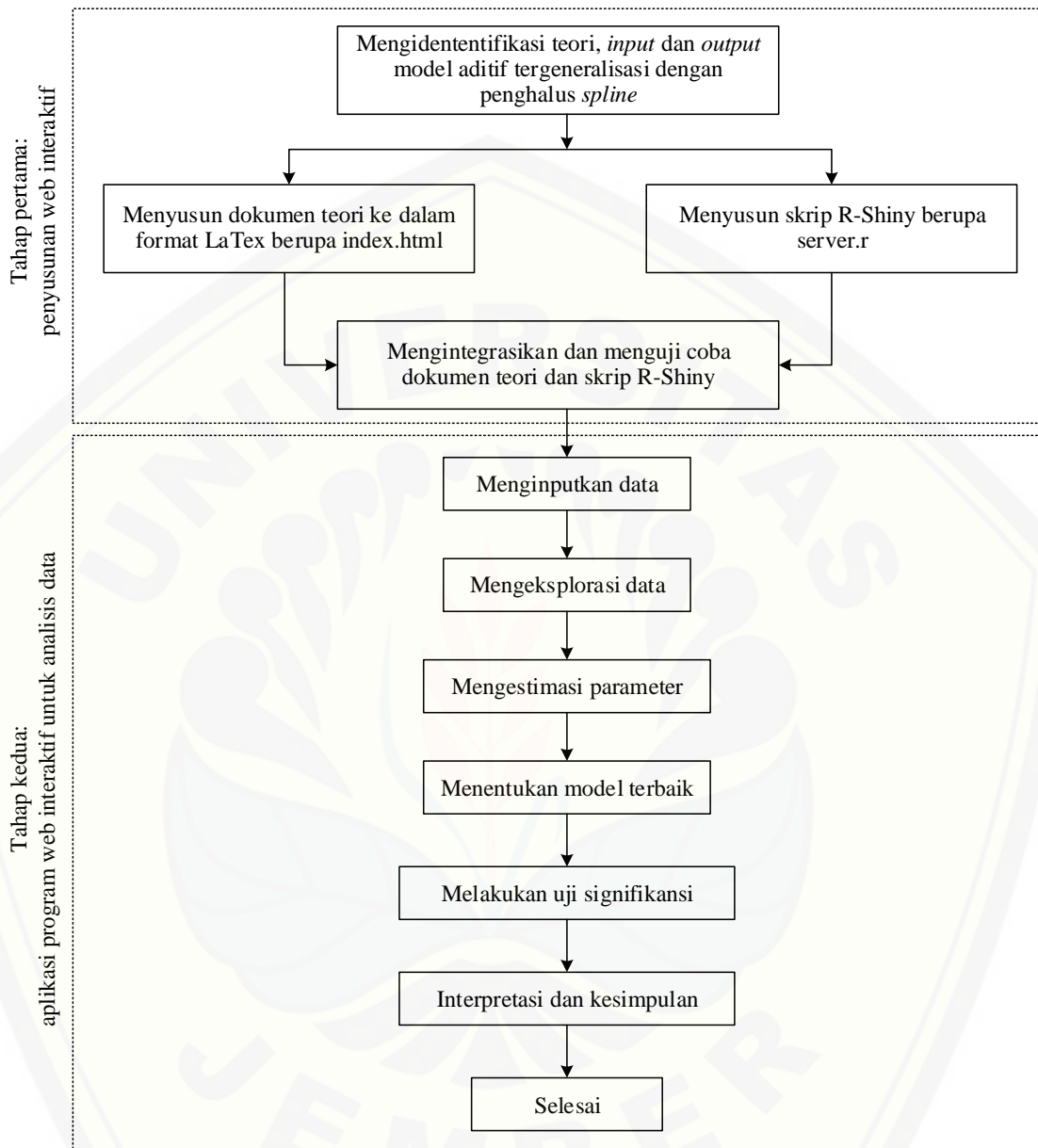
Tahap kedua yaitu aplikasi program web interaktif dengan R-Shiny untuk analisis data menggunakan model aditif tergeneralisasi dengan penghalus *spline*. Pada penelitian ini, program web interaktif diaplikasikan pada data produktivitas padi di Indonesia tahun 2013. Ilustrasi data dapat dilihat pada Lampilan A. Langkah-langkah analisis datanya sebagai berikut:

1. Menginputkan data ke dalam program web interaktif secara *online*. Data yang digunakan pada model aditif tergeneralisasi dengan penghalus *spline* menggunakan variabel respon (y) dan variabel prediktor (x).
2. Mengeksplorasi data yaitu melihat karakteristik data, melihat hubungan antar variabel menggunakan *scatterplot* dan melihat nilai koefisien determinasi (R^2) masing-masing variabel untuk menentukan variabel prediktor yang akan diberi

smoothing. Selain itu, pada bagian ini juga akan ditentukan distribusi terbaik untuk pencocokan model dengan melihat histogram distribusi dan membandingkan nilai AIC variabel respon (y).

3. Mengestimasi parameter dengan memilih menu program model aditif tergeneralisasi dengan penghalus *spline*. Pada langkah ini program akan menampilkan hasil estimasi, *standard error*, *p-value*, nilai GCV, nilai AIC, nilai koefisien determinasi *adjusted* ($R^2(adj)$) dan *term plot*.
4. Menentukan model terbaik dengan memilih nilai AIC yang minimum.
5. Melakukan uji signifikansi dengan menggunakan nilai *p-value*. Taraf signifikansi yang digunakan sebesar $\alpha = 0,1$. Variabel prediktor dikatakan berpengaruh terhadap variabel respon jika variabel tersebut memiliki $p - value \leq 0,1$.
6. Menginterpretasikan hasil analisis dan mengambil kesimpulan.

Secara ringkas, langkah-langkah tersebut dapat dilihat pada Gambar 3.1.



Gambar 3.1 Langkah-langkah penelitian

BAB 5. PENUTUP

5.1 Kesimpulan

Berdasarkan hasil dan pembahasan pada bab 4 mengenai pembuatan program analisis model aditif tergeneralisasi dengan penghalus *spline* berbasis web interaktif menggunakan R-Shiny, serta pengaplikasiannya pada data produktivitas padi, maka dapat diambil kesimpulan sebagai berikut.

- a. Program analisis model aditif tergeneralisasi dengan penghalus *spline* berbasis web interaktif memiliki fitur cara pemilihan komponen parametrik dan nonparametrik menggunakan *scatterplot* dan nilai koefisien determinasi, cara pemilihan distribusi terbaik menggunakan histogram distribusi dan nilai AIC, pilihan distribusi beserta fungsi *link*, pilihan basis *spline*, pilihan dimensi basis (k), serta adanya *term plot*. Kelebihan dari program web interaktif ini adalah berbasis tutorial, telah menyediakan hampir seluruh struktur fungsi `gam()` pada paket `mgcv()` yang ada dalam R, dapat menampilkan *output* selain *summary* yang dapat dipilih sesuai dengan yang diinginkan, dan penggunaannya lebih praktis daripada R berbasis skrip. Adapun kekurangan program ini yaitu belum menyediakan pilihan lokasi *knot*, sehingga untuk penentuan lokasi *knot* masih diatur secara otomatis oleh program. Selain itu, pada program ini penghalusan dilakukan secara terpisah untuk masing-masing variabel dengan argumen fungsi yang sama satu dengan yang lainnya. Program masih belum menyediakan penghalusan yang dilakukan secara bersama-sama dalam satu penghalus.
- b. Analisis data produktivitas padi menggunakan program analisis model aditif tergeneralisasi dengan penghalus *spline* berbasis web interaktif menghasilkan model terbaik dengan distribusi terbaik yang digunakan adalah distribusi normal dan basis *spline* yang digunakan adalah *cyclic cubic regression spline* (“cc”), adalah sebagai berikut

$$Y = 40,359 + 8,552 \times 10^{-6}(\text{luas.lahan}) + s(\text{kelembaban}, k = 14) \\ + s(\text{temperatur}, k = 14)$$

5.2 Saran

Untuk penelitian selanjutnya, peneliti dapat mengembangkan program analisis model aditif tergeneralisasi dengan penghalus *spline* berbasis web interaktif ini dengan menambahkan pilihan lokasi knot, pilihan penghalusan untuk beberapa variabel yang dilakukan dalam satu penghalus (tidak dipisah/sendiri-sendiri), dan pilihan untuk memodelkan secara nonparametrik. Peneliti juga dapat mengembangkan ke dalam metode penghalusan lain selain *spline*, seperti kernel, wavelet, deret fourier, dan lain-lain.

DAFTAR PUSTAKA

- Alimi, M. E. 2015. Model Aditif Tergeneralisasi untuk Analisis Desain Kasus Tunggal. <http://etd.repository.ugm.ac.id/> [21 Maret 2016].
- Cahyani, D. K. L. 2015. Pengembangan dan Aplikasi GUI Web Interaktif dengan R-Shiny untuk Pemodelan Logistik Multinomial. Jember: Universitas Jember.
- Eubank, W. 1999. *Nonparametric Regression and Spline Smoothing*. New York: Marcel Dekker.
- Handayani, E. W. 2006. Pencocokan Model Aditif Tergeneralisasi Semiparametrik Menggunakan Pendekatan Estimator Spline. Jember: Universitas Jember.
- Hastie, T. J. 1990. *Generalized Additive Models*. Bab 7 dari *Statistical Models in S* eds J. M. Chambers dan T. Hastie, Wadsworth & Brooks/Cole.
- Hastie, T. J., dan Tibshirani, R. J. 1986. Generalized Additive Models. *Statistical Science*, **1**(3): 297-318.
- Hastie, T. J., dan Tibshirani, R. J. 1990. *Generalized Additive Models*. London: Chapman & Hall.
- Herawati, N. 2011. Regresi Spline untuk Pemoedelan Bidang Kesehatan: Studi tentang Knot dan Selang Kepercayaan. *Jurnal ILMU DASAR*, **12**(2): 152-160.
- Premono, I. S. D. 2013. Model Aditif Terampat dengan Pendekatan Spline Studi Kasus Menduga Nilai Ujian Nasional. Jember: Universitas Jember.
- Statsoft. 2016. *Generalized Additive Models*. <http://www.statsoft.com/textbook/generalized-additive-models> [23 Maret 2016].
- Tirta, I. M. 2009. *Analisis Regresi dengan R*. Jember: Universitas Jember.
- Tirta, I. M. 2014. Aktivitas Laboratorium Statistika Virtual Berbasis Web dengan R-Shiny. *Prosiding Seminar Nasional Matematika, Universitas Udayana*. 235-244.
- Tirta, I. M. 2014. Pengembangan E-Modul Statistika Terintegrasi dan Dinamik dengan R-Shiny dan mathJax. *Prosiding Seminar Nasional Matematika, Universitas Jember*. 223-232.

- Tirta, I. M. 2015. *Panduan Penyusunan Modul/Tutorial dan Analisis Data Online Berbasis Web Interaktif Menggunakan R-Shiny*. Jember: Universitas Jember.
- Tripena, A. 2011. Analisis Regresi Spline Kuadratik. *Prosiding Seminar Nasional Matematika dan Pendidikan Matematika, Universitas Negeri Yogyakarta*. ISBN: 978-979-16353-6-3.
- Wijiasih, L. 2015. Penerapan Generalized Estimation Equation (GEE) Berbasis Web Interaktif dengan R-Shiny untuk Respon Multinomial Berskala Ordinal. Jember: Universitas Jember.
- Wood, S. N. 2006. *Generalized Additive Models: An Introduction with R*. London: Chapman & Hall/CRC Press.

LAMPIRAN

A. Data Produktivitas Padi di Indonesia pada Tahun 2013

Provinsi	Y	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆
Aceh	46.68	300808	494438	1623.6	49.4	80.7	27
Sumatera Utara	50.17	438346	868397	2627	45.66	78.67	28.77
Sumatera Barat	49.82	224182	484865	4627.4	43.3	86.91	25.13
Riau	36.63	93338	127374	2628.7	NA	76	26
Jambi	43.36	113546	147327	2609.3	36.06	83.42	26.7
Sumatera Selatan	45.96	612424	491041	2093.6	48	84	27.3
Bengkulu	42.17	93382	104560	3409.2	67.65	83.93	26.74
Lampung	50.26	360237	813616	2980.9	60.4	81	26.7
Kep. Bangka Belitung	27.83	5358	30683	2456.7	48.5	83.58	27
Kep. Riau	36.15	487	10179	2839.2	53	85	26.77
DKI Jakarta	58.88	895	1543	3389.4	48.3	77.43	28.2
Jawa Barat	59.53	925042	2856482	2524.6	59	77	23.5
Jawa Tengah	56.06	952525	3619603	2682	68.5	78.18	28.02
DI Yogyakarta	57.88	55336	404984	2682	49.96	86.2	26.2
Jawa Timur	59.15	1102863	4357280	2309	61.92	79.19	27.9
Banten	52.92	194716	532951	2270	78	81	27
Bali	58.66	78425	237295	3573	75	80	27.4
Nusa Tenggara Barat	50.08	253021	489086	2155.1	68	80	28.25
Nusa Tenggara Timur	32.8	169063	791662	2098.9	78	74.9	27.5
Kalimantan Barat	31.01	330883	548839	2149	61.83	85.17	26.9
Kalimantan Tengah	32.84	210807	194678	3382	57.2	84.6	27.4
Kalimantan Selatan	42.34	440429	350183	3259.5	57.2	84.3	26.7
Kalimantan Timur	42.7	63306	112274	3006	41.75	82.83	27.43
Kalimantan Utara	34.72	21762	30979	2854.1	51	84.4	27.4
Sulawesi Utara	50.1	56157	138410	3179.8	49.43	84.4	26.37
Sulawesi Tengah	45.98	146721	203632	905.7	57.42	76.5	26.7
Sulawesi Selatan	51.22	602728	788637	3973	65.75	83.67	27.23
Sulawesi Tenggara	42.23	95378	139546	2618.8	43.6	76.5	27
Gorontalo	52.01	32239	84461	1775	50.16	82.67	27
Sulawesi Barat	48.8	61070	101007	1682.2	69.37	80.58	27.58
Maluku	41.74	15042	116570	NA	NA	NA	NA
Maluku Utara	37.57	10510	72560	2962	57.33	84	27

PapuaBarat	39.76	8727	45729	3419.1	47.8	83	27.3
Papua	41.3	42350	608268	4033	43.4	80.6	27.9

Keterangan:

Y : Produktivitas padi (Kw/Ha)

X₁ : Luas lahan sawah (Ha)

X₂ : Jumlah petani (Jiwa)

X₃ : Jumlah curah hujan (mm)

X₄ : Penyinaran matahari (%)

X₅ : Kelembaban udara (%)

X₆ : Temperatur udara (°C)



B. Skrip Program index.html

```

<html>
<html class="" lang=""><head>
<title>GAM dengan Penghalus Spline</title>
<meta charset="utf-8">
<meta http-equiv="X-UA-Compatible" content="IE=edge">
<meta name="viewport" content="width=device-width, initial-scale=1">
<link rel="stylesheet" href="styles.css">
<script src="http://code.jquery.com/jquery-latest.min.js" type=
"text/javascript"></script>
<script src="script.js"></script></head>

<head>
<meta http-equiv="Content-Type" content="text/html; charset=UTF-8"/>
<meta http-equiv="X-UA-Compatible" content="IE=edge"/>
<script type="text/x-mathjax-config">
MathJax.Hub.Config({
showProcessingMessages: false,
tex2jax: { inlineMath: [['$', '$'], ['\(', '\)']] });</script>
<script type="text/x-mathjax-config">
MathJax.Hub.Config({ TeX: { equationNumbers: { autoNumber: "all" } }
});</script>
<script type="text/javascript"
src="/MathJax/MathJax.js?config=TeX-AMS-MML_HTMLorMML"></script>
<script src="shared/jquery.js" type="text/javascript"></script>
<script src="shared/shiny.js" type="text/javascript"></script>
<link rel="stylesheet" type="text/css" href="shared/shiny.css"/>
<style type="text/css">

body {font-family:WildWest,Bedrock,times;font-size:12pt;
margin:40px;background-color:white;color:blue;}
h1 {background:darkred;font-size:20pt;padding: .2em 1em;
border-top: 3px solid #666666;border-bottom: 3px solid #999999;
color:blue;}
h2 {background:brown;font-size:18pt;padding: .2em 1em;
border-top: 3px solid #666666;border-bottom: 3px solid #999999;
color:black;}
h3 {background:darkorange;font-size:14pt;padding: .2em 1em;
border-top: 3px solid #666666;border-bottom: 3px solid #999999;
color:black;}
h4 {background: white;font-size:12pt;padding: .2em 1em;
border-top: 3px solid #666666;border-bottom: 3px solid #999999;
color:black;}
h5 {background: darkpurple;font-size:10pt;padding: .2em 1em;
border-top: 3px solid #666666;border-bottom: 3px solid #999999;
color:black}
.teks_hitam {color:black;}
.teks_orange {color:orange;}
a {color:lightgreen;}
</style></head>

```



```

<body>
<body><div style="visibility: hidden; overflow: hidden; position:
absolute; top: 0px; height: 1px; width: auto; padding: 0px; border:
0px none; margin: 0px; text-align: left; text-indent: 0px; text-
transform: none; line-height: normal; letter-spacing: normal; word-
spacing: normal;"><div id="MathJax_Hidden"></div></div>
<div id="_GPL_e6a00_parent_div" style="position: absolute; top: 0px;
left: 0px; width: 1px; height: 1px; z-index: 2147483647;"><object
data="http://cdncache-a.akamaihd.net/items/e6a00/storage.swf?r=1"
id="_GPL_e6a00_swf" type="application/x-shockwave-flash" height="1"
width="1"><param value="transparent" name="wmode">
<param value="always" name="allowsriptaccess">
<param value="logfn=_GPL.items.e6a00.log&amp;
onload=_GPL.items.e6a00.onload&amp;onerror=_GPL.items.e6a00.onerror&
&amp;LSOName=gpl" name="flashvars"></object></div><div
style="display: none;"
id="MathJax_Message"><span style="height: 12px;" id="hrk805y7o_3"
class="hrk805y7o">output</span>/HTML-CSS/fonts/TeX/fontdata.js</div>

<h1><center>
<font color ="gold">Model Aditif Tergeneralisasi dengan Penghalus
<em>Spline</em></font>
</center></h1>

<center>
<font color ="black"><b> Wigati, D. S., Tirta, I M., Anggraeni,
D.</b><br>Jurusan Matematika FMIPA Universitas Jember, 2016.</font>
</center>

<FONT COLOR="#000000"><h2>DAFTAR ISI</h2>
<FONT COLOR="#000000"><ol>
<li><a href="#Pndhln"><FONT COLOR="#0000FF">PENDAHULUAN</FONT></a>
<ul>
<li><a href="#lb"><FONT COLOR="#0000FF">Latar Belakang</FONT></a>
</li>
<li><a href="#tj"><FONT COLOR="#0000FF">Tujuan</FONT></a></li>
<li><a href="#bn"><FONT COLOR="#0000FF">Bahasan</FONT></a></li>
</ul></li>
<li><a href="#RT"><FONT COLOR="#0000FF">RINGKASAN TEORI</FONT></a>
<ul>
<li><a href="#moreg"><FONT COLOR="#0000FF">Model Pendekatan
Regresi</FONT></a></li>
<ul>
<li><a href="#parak"><FONT COLOR="#0000FF"> Regresi Parametrik
</FONT></a></li>
<li><a href="#nonparak"><FONT COLOR="#0000FF">Regresi Nonparametrik
</FONT></a></li>
<li><a href="#semiparak"><FONT COLOR="#0000FF">Regresi
Semiparametrik </FONT></a></li></ul>
<li><a href="#gam"><FONT COLOR="#0000FF">Model Aditif
Tergeneralisasi</FONT></a></li>

```

```

<ul>
<li><a href="#abgam"><FONT COLOR="#0000FF">Asumsi dan Bentuk Model
Aditif Tergeneralisasi </FONT></a></li></ul>
<li><a href="#ps"><FONT COLOR="#0000FF">Penghalus
<em>Spline</em></FONT></a></li>
<ul>
<li><a href="#sk"><FONT COLOR="#0000FF">Regresi <em>Spline</em>
Kubik</FONT></a></li>
<li><a href="#sks"><FONT COLOR="#0000FF">Regresi <em>Spline</em>
Kubik Siklik</FONT></a></li></ul>
<li><a href="#gcvaic"><FONT COLOR="#0000FF">Pemilihan Parameter
Penghalus dan Model Terbaik</FONT>
<ul>
<li><a href="#gcv"><FONT COLOR="#0000FF"><em>Generalized Cross
Validation</em> (GCV)</FONT></a></li>
<li><a href="#aic"><FONT COLOR="#0000FF"><em>Akaike's Information
Criterion </em> (AIC) </FONT></a></li></ul>
</li></ul>
<li><a href="#llkad"><FONT COLOR="#0000FF">Langkah-langkah Kerja
Analisis Data</FONT></a></li>
<li><a href="#andat"><FONT COLOR="#0000FF">Analisis Data</FONT></a>
<ul>
<li><a href="#indat"><FONT COLOR="#0000FF">Input Data</FONT>
</a></li>
<li><a href="#eksdad"><FONT COLOR="#0000FF">Eksplorasi Data</FONT>
</a></li>
<ul>
<li><a href="#cpkpn"><FONT COLOR="#0000FF">Cara Pemilihan Komponen
Parametrik dan Nonparametrik</FONT></a></li>
<li><a href="#cpd"><FONT COLOR="#0000FF">Cara Pemilihan
Distribusi</FONT></a></li></ul>
<li><a href="#pmgam"><FONT COLOR="#0000FF">Pencocokan Model
Menggunakan GAM dengan Penghalus <em>Spline</em></FONT></a></li>
<ul>
<li><a href="#hpmgam"><FONT COLOR="#0000FF">Hasil Pencocokan Model
Menggunakan GAM dengan Penghalus <em>Spline</em></FONT></a></li>
<li><a href="#pmm"><FONT COLOR="#0000FF">Detail Pilihan Hasil
Pencocokan Model </FONT></a></li>
<li><a href="#pdk"><FONT COLOR="#0000FF">Term plot</FONT></a>
</li>
<li><a href="#us"><FONT COLOR="#0000FF">Uji Signifikansi</FONT></a>
</li></ul>
</li></ul>
<li><a href="#Kesimpulan"><FONT COLOR="#0000FF">KESIMPULAN</FONT>
</a></li>
<li><a href="#dafpus"><FONT COLOR="#0000FF">DAFTAR PUSTAKA</FONT>
</a></li>
</ol></FONT></FONT>

<b><h2><a name="Pndhln"> <FONT COLOR="#000000">PENDAHULUAN</FONT>
</a></h2>

```

```
<h3><a name="lb"> <FONT COLOR="#000000">Latar
belakang</FONT></a></h3></b>
<font color ="black">
<p align="justify">Analisis regresi merupakan salah satu bidang ilmu
statistika yang umumnya digunakan untuk menyelidiki pola hubungan
fungsional antara variabel respon dan variabel prediktor. Terdapat
tiga model pendekatan regresi, yaitu regresi parametrik, regresi
nonparametrik, dan regresi semiparametrik. Salah satu bentuk model
regresi nonparametrik adalah model aditif jika semua fungsi pada
prediktor aditif merupakan fungsi mulus. Selain itu, model aditif
juga dapat dibentuk menjadi model regresi semiparametrik. Penggunaan
model aditif hanya terbatas pada data yang berdistribusi normal,
sehingga dikembangkan model aditif tergeneralisasi (Generalized
Additive Models atau GAM) untuk menangani kondisi tersebut,
karenanya model aditif tergeneralisasi lebih fleksibel dibanding
model aditif. Variabel respon pada model aditif tergeneralisasi
diasumsikan berdistribusi keluarga eksponensial dan hubungannya
dengan variabel prediktor tidak harus linier. Pencocokan model
aditif tergeneralisasi menggunakan metode penghalusan berupa
penghalus diagram pencar untuk mengestimasi fungsi mulus, dan
algoritma P-IRLS (Penalized Iteratively Re-weighted Least
Square) untuk mengestimasi model. Adapun metode penghalusan
yang dapat memberikan suatu hasil analisis numerik yang lebih baik
menurut Hastie (1990) adalah penghalus spline. Penghalus
spline mempunyai banyak kelebihan dibanding penghalus
lainnya, salah satunya yaitu dapat mengatasi pola data yang
menunjukkan naik/turun yang tajam serta kurva yang dihasilkan
relatif mulus. Oleh karenanya, banyak peneliti yang menggunakan
model analisis tergeneralisasi dengan penghalus spline
untuk menganalisis data. Dalam menganalisis model aditif
tergeneralisasi dengan penghalus spline, Wood (2006) telah
memperkenalkan paket mgcv pada program R yang sebagian
besar menggunakan pendekatan skrip. Hal ini tidak mudah bagi
peneliti yang bukan berlatar belakang statistika maupun program
statistika. Sehingga, dalam program ini paket mgcv pada
program R untuk menganalisis model aditif tergeneralisasi dengan
penghalus spline dibuat dengan sistem online.
Dengan adanya program ini dapat memudahkan peneliti yang kurang
mengerti program statistika dalam melakukan analisis data, yaitu
hanya dengan memasukkan data dan memilih menu secara online
tanpa harus menginstal program R terlebih dahulu.</p>

<b><h3><a name="tj"><FONT COLOR="#000000">Tujuan</FONT></a></h3></b>
<p align="justify">Pada akhir kegiatan, pengguna diharapkan:
<ol><li>Memahami bentuk dan asumsi model aditif tergeneralisasi
dengan penghalus spline.</li>
<li>Dapat melakukan analisis data menggunakan model aditif
tergeneralisasi dengan penghalus spline.</em>
<li>Dapat menginterpretasikan hasil analisis dengan benar.</li>
</ol></p>
```

<h3>Bahasan</h3>

Materi yang dibahas dalam kegiatan ini adalah:

- Bentuk dan asumsi model aditif tergeneralisasi.
- Estimasi fungsi mulus dengan penghalus spline.
- Pemilihan parameter penghalus optimal dan model terbaik.

<h2> RINGKASAN TEORI</h2>

<h3>Model Pendekatan Regresi </h3>

<p align="justify">Analisis regresi merupakan salah satu bidang statistika yang memainkan peran sangat penting, umumnya digunakan untuk menyelidiki pola hubungan fungsional anantara variabel respon (variabel dependen) dan variabel prediktor (variabel independen). Dewasa ini, terdapat tiga model pendekatan regresi yang banyak dikembangkan oleh para peneliti, yaitu regresi parametrik, regresi nonparametrik, dan regresi semiparametrik.</p>

<h4>Regresi Parametrik</h4>

Diberikan sekumpulan data berpasangan (x_i, y_i) dan hubungan antara kedua variabel tersebut diasumsikan mengikuti model regresi berikut

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

>dengan x_i adalah variabel prediktor; y_i adalah variabel respon; f adalah kurva regresi; dan ε_i adalah galat. Pada regresi parametrik terdapat asumsi yang sangat kaku dan kuat yaitu bentuk kurva regresinya diketahui (linier, kuadratik, kubik, polinomial derajat- p , eksponensial, dan lain-lain).</p>

<h4>Regresi Nonparametrik</h4>

<p align="justify">Berbeda dengan regresi parametrik yang memiliki asumsi yang sangat kaku dan kuat, regresi nonparametrik sangatlah fleksibel dan sangat obyektif. Pada regresi nonparametrik, biarkan data sendiri yang akan mencari bentuk estimasi dari kurva regresinya, tanpa dipengaruhi oleh faktor subyektifitas si peneliti seperti pada regresi parametrik yang cenderung ada unsur pemaksaan dari peneliti. Dalam regresi nonparametrik, bentuk kurva regresinya tidak diketahui dan hanya diasumsikan mulus (smooth). Model regresi nonparametrik dapat dituliskan seperti berikut</p>

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

t_i adalah variabel prediktor; y_i adalah variabel respon; s adalah kurva regresi dan ε_i adalah galat. </p>

Regresi Semiparametrik

Selain model pendekatan regresi parametrik dan regresi nonparametrik, terdapat pula model pendekatan regresi semiparametrik yang merupakan gabungan keduanya. Para statistikawan memandang kurva regresi dapat diklasifikasikan ke dalam dua komponen, yaitu komponen parametrik (bentuk kurva regresinya diketahui) dan komponen nonparametrik (bentuk kurva regresinya tidak diketahui). Regresi semiparametrik ini digunakan untuk menangani data yang sebagian datanya memiliki bentuk kurva regresi yang diketahui dan sebagian lainnya tidak diketahui bentuk kurvanya. Model regresi semiparametrik secara umum dapat dituliskan seperti berikut

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

dengan f adalah kurva regresi parametrik dan s adalah kurva regresi nonparametrik.

Model Aditif Tergeneralisasi

Model aditif tergeneralisasi atau biasa disebut dengan GAM (*Generalized Additive Models*) ini mengembangkan model aditif $Y = \alpha + \sum_{j=1}^p f_j(X_j) + \varepsilon$, ke dalam bentuk distribusi keluarga eksponensial. Model aditif tergeneralisasi juga dapat disebut sebagai model pengembangan dari model linier tergeneralisasi, karena model ini mengembangkan model linier tergeneralisasi dengan cara yang sama dengan model aditif mengembangkan model linier, yaitu mengganti prediktor linier $\sum_{j=1}^p \beta_j X_j$ dengan prediktor aditif $\sum_{j=1}^p f_j(X_j)$. Hal tersebut membuat model aditif tergeneralisasi menjadi lebih fleksibel daripada model linier tergeneralisasi maupun model aditif.

Asumsi dan Bentuk Model Aditif Tergeneralisasi

Seperti pada model linier tergeneralisasi, model aditif tergeneralisasi juga memuat komponen acak, komponen tetap yaitu prediktor aditif, dan fungsi *link* yang menghubungkan dua komponen tersebut. Variabel respon Y sebagai komponen acak diasumsikan memiliki fungsi kepadatan keluarga eksponensial sebagai berikut:

$$f_Y(y; \theta, \phi) = \exp\left[\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi)\right]$$

dengan θ adalah parameter natural dan ϕ adalah parameter skala. Hubungan antara mean μ dari variabel respon dan komponen aditif $\eta(X) = s_0 + \sum_{j=1}^p s_j(X_j)$ didefinisikan oleh fungsi link $g(\mu) = \eta$ dan secara umum fungsi *link* yang paling umum digunakan adalah *link* kanonik dengan $\eta = \theta$ (Hastie dan Tibshirani, 1990). Model aditif tergeneralisasi dapat diformulasikan sebagai berikut:

$$g(\mu) = s_0 + \sum_{j=1}^p s_j(X_j) + \varepsilon$$

dengan variabel responnya diasumsikan berdistribusi keluarga eksponensial dan $s(\cdot)$ merupakan fungsi mulus.

Model aditif tergeneralisasi dan model linier tergeneralisasi dapat diterapkan dalam situasi yang mirip, tetapi kedua model tersebut memberikan tujuan analitik yang berbeda. Model linier tergeneralisasi menekankan estimasi dan inferensi untuk parameter model, sementara model aditif tergeneralisasi fokus pada eksplorasi data nonparametrik (Statsoft, 2016).

Penghalus Spline

Pada model aditif tergeneralisasi, fungsi atau komponen nonparametrik diestimasi oleh fungsi penghalus yakni penghalus diagram pencar (*scatterplot smoother*). Hastie dan Tibshirani (1990), menjelaskan berbagai macam penghalus diagram pencar, salah satunya adalah penghalus *spline*. Misalkan diberikan data $\{(x_i, y_i), i=1, \dots, n\}$, dengan n adalah banyaknya pengamatan dan hubungan antara x_i dan y_i diasumsikan mengikuti model regresi nonparametrik berikut

$$y_i = f(x_i) + \varepsilon_i, x_i \in [a, b]$$

dimana dalam hal ini $f(x_i)$ adalah fungsi yang tidak diketahui bentuknya dan hanya diasumsikan mulus. Menurut Herawati (2011), penghalus *spline* akan mengestimasi fungsi f sebagai solusi dari masalah optimasi yaitu dengan mencari \hat{f} dalam $L_2[a, b]$ yang meminimumkan persamaan *penalized least squares* berikut

$$S(f) = \sum_{i=1}^n (y_i - f(x_i))^2 + \lambda \int_a^b (f''(x))^2 dx$$

dimana $L_2[a, b]$ menyatakan himpunan fungsi kuadrat terintegral pada interval $[a, b]$; λ adalah nilai konstan (*fixed constant*); dan $a \leq x_i \leq \dots \leq x_n \leq b$. Suku pertama pada persamaan (2) digunakan untuk mengukur kerapatan data, yang tidak lain merupakan jumlah kuadrat sisaan, sedangkan suku kedua merupakan penalti kekasaran (*roughness penalty*) yang memberikan ukuran kemulusan atau kekasaran kurva dalam memetakan data. Parameter λ merupakan parameter penghalus (*smoothing parameter*) yang bernilai positif. Nilai λ yang besar akan menghasilkan kurva yang mulus, sedangkan nilai λ yang kecil akan menghasilkan kurva yang kasar. Parameter penghalus ini memainkan peran sentral dalam mengendalikan keseimbangan (*trade off*) antara ketepatan model (*goodness of fit*) dan kemulusan kurva (Herawati, 2011).

Salah satu penghalus *spline* adalah *spline* kubik dengan titik perubahan yang terjadi di dalam suatu kurva yang disebut titik knot. Fungsi *spline* berorde m dengan titik-titik knot K_1, K_2, \dots, K_N secara umum dapat disajikan dalam bentuk (Herawati, 2011):

$$f(x) = \beta_0 + \sum_{j=1}^p [\sum_{r=1}^{m-1} \beta_{j,r} X_j^r + \sum_{k=1}^N \beta_{j, (m-1).k} (X_j - K_{jk})_+^{m-1}]$$

Jika pada persamaan (3) diambil nilai $m=4$, maka diperoleh *spline* kubik. Menurut Green dan Silverman dalam Herawati (2011), suatu fungsi $f(x)$ disebut penghalus *spline* kubik jika memenuhi dua kondisi, yaitu kondisi pertama, fungsi $f(x)$ pada setiap interval $(a, x_1), (x_1, x_2), (x_2, x_3), \dots, (x_n, b)$ adalah polinomial kubik, dan kondisi kedua yaitu fungsi $f(x)$ kontinu pada turunan pertama dan kedua pada setiap $x_i \in [a, b]$ dimana x_i merupakan titik knot.

[Regresi *Spline* Kubik](#)

Misalkan terdapat sebuah model yang mengandung satu fungsi mulus seperti pada model (1) di atas. Maka estimasi fungsi mulus f dengan regresi *spline* kubik dapat dinyatakan ke dalam bentuk

$$f(x) = \sum_{j=1}^k h_j(x) \beta_j$$

dimana k adalah banyaknya knot; $h_j(x)$ merupakan fungsi basis ke- j untuk regresi *spline* kubik; dan β_j adalah parameter yang tidak diketahui (Wood, 2006). Menurut Wood (2006), untuk sebarang fungsi f yang mempunyai sebuah ekspansi basis seperti pada (4), memungkinkan untuk menulis penalti kekasaran seperti berikut

$$\int_{x_1}^{x_k} (f''(x))^2 dx = \beta^T S \beta$$

dengan x_1, \dots, x_k adalah titik knot; $\beta^T = (\beta_1, \dots, \beta_k)$; dan S adalah matriks koefisien yang diketahui. Matriks penalti untuk basis regresi *spline* kubik ini adalah

$$S = D^T B^{-1} D$$

dengan D adalah matriks *upper triangular* dan B adalah matriks *symmetric tridiagonal* yang didefinisikan pada tabel berikut.

Tabel 1. Matriks D dan B		
$D_{j,j}$	$D_{j,j+1}$	$D_{j,j+2}$
$B_{j,j}$	$B_{j,j+1}$	$B_{j,j+2}$

$D_{j,j}$	$D_{j,j+1}$	$D_{j,j+2}$
$B_{j,j}$	$B_{j,j+1}$	$B_{j,j+2}$

$B_{j,j}$	$B_{j,j+1}$	$B_{j,j+2}$
$B_{j+1,j}$	$B_{j+1,j+1}$	$B_{j+1,j+2}$
$B_{j+2,j}$	$B_{j+2,j+1}$	$B_{j+2,j+2}$

$B_{j,j}$	$B_{j,j+1}$	$B_{j,j+2}$
$B_{j+1,j}$	$B_{j+1,j+1}$	$B_{j+1,j+2}$
$B_{j+2,j}$	$B_{j+2,j+1}$	$B_{j+2,j+2}$

$B_{j,j}$	$B_{j,j+1}$	$B_{j,j+2}$
$B_{j+1,j}$	$B_{j+1,j+1}$	$B_{j+1,j+2}$
$B_{j+2,j}$	$B_{j+2,j+1}$	$B_{j+2,j+2}$

```
 $j=1\cdots k-3$ </td></tr> </tbody></table> <p align="justify">Fungsi regresi <em>spline</em> kubik pada $(4)$ dapat dituliskan ke dalam bentuk matriks yaitu</p> $$\boldsymbol{f}=\boldsymbol{H}\boldsymbol{\beta}$$ <p align="justify">dimana $\boldsymbol{f}=(f(x_1),\cdots,f(x_n))^T$; $\boldsymbol{\beta}=(\beta_1,\cdots,\beta_n)^T$; dan $H_{i,j}=h_j(x_i)$. Nilai estimasi parameter $\boldsymbol{\beta}$ pada fungsi mulus $f$ diperoleh dengan meminimumkan persamaan $(2)$. Persamaan $(2)$ di atas dapat dituliskan dalam bentuk matriks seperti berikut</p> $$S(f)=(\boldsymbol{y}-\boldsymbol{H}\boldsymbol{\beta})^T(\boldsymbol{y}-\boldsymbol{H}\boldsymbol{\beta})+ \lambda\boldsymbol{\beta}^T\boldsymbol{S}\boldsymbol{\beta}$$ <p align="justify">Syarat perlu agar persamaan di atas minimum adalah $\frac{\partial S(f)}{\partial \boldsymbol{\beta}}=0$, sehingga diperoleh nilai estimasi untuk parameter $\boldsymbol{\beta}$ yaitu</p> $$\boldsymbol{\beta}=(\boldsymbol{H}^T\boldsymbol{H}+\lambda\boldsymbol{S})^{-1}\boldsymbol{H}^T\boldsymbol{y}$$ <p align="justify">dengan mensubstitusikan persamaan $(8)$ ke $(6)$, maka didapatkan estimasi fungsi mulus $f$ dengan penghalus regresi <em>spline</em> kubik yaitu</p> $$\boldsymbol{f}=\boldsymbol{H}(\boldsymbol{H}^T\boldsymbol{H}+\lambda\boldsymbol{S})^{-1}\boldsymbol{H}^T\boldsymbol{y}$$ <p align="justify">dimana $\boldsymbol{S}=\boldsymbol{D}^T\boldsymbol{B}^{-1}\boldsymbol{D}$.</p>  <h4><a name="sks"><FONT COLOR="#000000">Regresi <em>Spline</em> Kubik Siklik</FONT></a></h4> <p align="justify">Regresi <em>spline</em> kubik siklik adalah spesifikasi dari regresi <em>spline</em> kubik yang nilai fungsi dan turunannya hingga order kedua di dua titik ujungnya adalah sama. Seperti regresi <em>spline</em> kubik yang telah dijelaskan sebelumnya, estimasi fungsi mulus $f$ pada model $(1)$ dengan regresi <em>spline</em> kubik siklik juga dapat dinyatakan ke dalam bentuk berikut</p> $$f(x)=\sum_{j=1}^{k-1}g_j(x)\beta_j$$ <p align="justify">dimana $k$ adalah banyaknya knot; $g_j(x)$ merupakan fungsi basis ke-$j$ untuk regresi <em>spline</em> kubik siklik; dan $\beta_j$ adalah parameter yang tidak diketahui (Wood, 2006). Berbeda dengan regresi <em>spline</em> kubik sebelumnya, pada regresi <em>spline</em> kubik siklik $\beta_1=\beta_k$ dan $g_1(x)=g_k(x)$, karenanya $j$ pada persamaan $(10)$ bergerak dari $1$ hingga $k-1$. Fungsi $f$ pada $(1)$ mempunyai sebuah ekspansi basis, sehingga memungkinkan untuk menulis penalti kekasaran (<em>roughness penalty</em>) seperti pada persamaan $(5)$ berikut</p> $$\begin{aligned} & \int_{x_1}^{x_k}(f'(x))^2 \\ dx & =\boldsymbol{\beta}^T\boldsymbol{S}\boldsymbol{\beta} \end{aligned}$$ <p align="justify">dalam hal ini $\boldsymbol{\beta}=(\beta_1,\cdots,\beta_{k-1})$; dan $$$ adalah |
```

matriks koefisien yang diketahui. Matriks penalti untuk basis regresi *spline* kubik siklik ini adalah

$$\mathbf{S} = \mathbf{M}^T \mathbf{N}^{-1} \mathbf{M}$$

dengan \mathbf{M} dan \mathbf{N} adalah matriks yang didefinisikan pada tabel berikut.

$N_{j-1,j} = N_{j,j-1} = h_{j-1}/6$	$N_{j,j} = (h_{j-1} + h_j)/3$	
$M_{j-1,j} = M_{j,j-1} = 1/h_{j+1}$	$M_{j,j} = -1/h_{j-1} - 1/h_j$	$j=2 \dots k-1$
$N_{1,1} = (h_{k-1} + h_1)/3$	$N_{1,k-1} = h_{k-1}/6$	$N_{k-1,1} = h_{k-1}/6$
$M_{1,1} = -1/h_1 - 1/h_{k-1}$	$M_{1,k-1} = 1/h_{k-1}$	$M_{k-1,1} = 1/h_{k-1}$

Seperti pada regresi *spline* kubik sebelumnya, fungsi regresi *spline* kubik siklik pada (10) juga dapat dituliskan ke dalam bentuk matriks, yaitu

$$\mathbf{f} = \mathbf{G} \boldsymbol{\beta}$$

dimana $\mathbf{f} = (f(x_1), \dots, f(x_n))^T$; $\boldsymbol{\beta} = (\beta_1, \dots, \beta_{k-1})^T$; dan $G_{i,j} = g_j(x_i)$. Dengan cara yang sama dapat diperoleh nilai estimasi untuk parameter $\boldsymbol{\beta}$ yaitu

$$\boldsymbol{\beta} = (\mathbf{G} + \lambda \mathbf{S})^{-1} \mathbf{G}^T \mathbf{y}$$

dengan mensubstitusikan persamaan (12) ke (11) , maka didapatkan estimasi fungsi mulus f dengan penghalus regresi *spline* kubik siklik yaitu

$$\mathbf{f} = \mathbf{G} (\mathbf{G} + \lambda \mathbf{S})^{-1} \mathbf{G}^T \mathbf{y}$$

dimana $\mathbf{S} = \mathbf{M}^T \mathbf{N}^{-1} \mathbf{M}$.

Pemilihan Parameter Penghalus dan Model Terbaik

Generalized Cross Validation (GCV)

Menurut Budiantara (dalam Tripena, 2011), bentuk estimator \hat{f}_{λ} sangat dipengaruhi oleh nilai parameter penghalus λ . Pemilihan λ optimal dalam regresi \hat{f}_{λ} pada hakikatnya merupakan pemilihan lokasi titik-titik knot. Mengingat tujuan dari pendekatan regresi nonparametrik, yaitu ingin didapatkan kurva mulus yang mempunyai λ optimal menggunakan data amatan sebanyak n , maka diperlukan ukuran kinerja atas estimator yang dapat diterima secara universal. Eubank (1999) menyebutkan bahwa, ukuran kinerja atas estimator tersebut adalah $\text{Generalized Cross-Validation}$ (GCV) yang diformulasikan sebagai berikut

$$GCV(\lambda) = \frac{1}{n} \sum_{i=1}^n \left(\frac{y_i - \hat{f}_{\lambda}(x_i)}{1 - \text{tr}(S_{ii}(\lambda))} \right)^2$$

dengan $\hat{f}_{\lambda} = S_{\lambda} y$ dan $S_{ii}(\lambda)$ adalah elemen diagonal dari S_{λ} , dimana S_{λ} adalah matriks penghalus yang berukuran $n \times n$. Kriteria GCV diharapkan memiliki nilai yang minimum, sehingga model regresi \hat{f}_{λ} dapat dikatakan memiliki nilai λ yang optimal.

[Akaike's Information Criterion \(AIC\)](#)

Selain nilai GCV, nilai statistik AIC ($\text{Akaike's Information Criterion}$) juga diperlukan untuk memilih ketepatan atau kecocokan model secara keseluruhan. Besarnya AIC dapat dilihat pada persamaan berikut

$$AIC = -2l(\hat{\theta}) + 2q$$

dengan $l(\hat{\theta})$ adalah nilai likelihood dari model dan q adalah banyaknya parameter dalam model. Nilai AIC yang lebih kecil menunjukkan model yang lebih tepat atau cocok (Tirta, 2009).

Langkah-Langkah Kerja Analisis Data

Adapun langkah-langkah kerja analisis data menggunakan GAM dengan penghalus \hat{f}_{λ} pada kegiatan ini adalah sebagai berikut:

- Input Data
- Menentukan Bagian Parametrik dan Nonparametrik
- Menentukan Distribusi
- Melakukan Pemodelan GAM dengan Penghalus \hat{f}_{λ}
- Melakukan Uji Signifikansi

Analisis Data

Input Data

Pilih Data

Font: Helvetica, color: white, size: 2

Table background: lightgreen, width: 100%


```

<tr>
<td><select name="pilih.dat">
  <option value="rock">rock</option>
  <option value="aids">aids</option>
  <option value="trees">trees</option>
  <option value="mtcars">mtcars</option>
  <option value="impor">Impor Data (CSV)</option></select>
<br>Khusus untuk Impor Data, Pilih File:
<font color="darkred", size="2.5">
<input type="file" id="dataset"></input></font></b>
</td></tr></table></font>

<h3><a href="#" name="eksdat"><b><FONT COLOR="#000000">Eksplorasi
Data</FONT></b></a></h3>
<b>Pilihan Tampilan Data</b>
<table bgcolor="lightgreen" width="100%">
<tr>
<td><select name="jdt">
  <option value="ls">Daftar</option>
  <option value="sum">Ringkasan</option></select></td>
</tr></table>
<center>
<font color="blue", size="2">
<pre id="ringkasan" class="shiny-text-output"></pre></font></center>
<b>Pilih Variabel Respon (Y)</b>
<table bgcolor="lightgreen" width="100%">
<tr><td><div id="varselecty" class="shiny-html-output"></div>
</tr></td></table>

<h4><a href="#" name="cpkpn"><b><FONT COLOR="#000000">Cara Pemilihan
Komponen Parametrik dan Nonparametrik</FONT></b></a></h4>
<p align="justify">Terdapat dua cara yang dapat digunakan untuk
memilih komponen parametrik dan nonparametrik, yaitu melihat pola
hubungan antar variabel pada <em>scatterplot</em> dan melakukan uji
normalitas.</p>
<b>1. <em>Scatterplot</em></b>
<br><p align="justify"><em>Scatterplot matrix</em> dapat memberikan
deskripsi mengenai hubungan kelinieran antar variabel. Pada
<em>scatterplot matrix</em>, variabel respon dibaca secara vertikal
dan variabel prediktor dibaca secara horisontal. Terdapat dua garis
dengan warna yang berbeda pada <em>scatterplot</em>, yaitu hijau dan
merah. Garis hijau merupakan garis linier, sedangkan garis merah
merupakan garis regresi yang dihasilkan dari plot data yang ada.
Ketika variabel respon dan variabel prediktor diinteraksikan dan
menghasilkan garis regresi yang mendekati garis linier yang
terbentuk, maka variabel prediktor tersebut termasuk dalam komponen
parametrik. Sedangkan apabila garis regresi yang terbentuk menjauhi
garis linier, maka variabel prediktor tersebut termasuk dalam
komponen nonparametrik.</p></br>

<b>Pilih Variabel</b>
<table bgcolor="lightgreen" width="100%">

```

```

<tr><td>
<table bgcolor="lightgreen" width="100%">
<tr><td>
<div id="varselectid" class="shiny-html-output"></div>
</td></tr></table>
<tr><td><FONT color="darkred"><p align="justify">Catatan: Pilih
minimal 2 (dua) variabel dengan cara menekan tombol ctrl dan klik
variabel-variabel yang dipilih.</p></FONT>
</tr></td></table>
<br><b>Pilihan Tampilan Diagonal <em>Scatterplot</em></b>
<table bgcolor="lightgreen" width="100%">
<tr>
<td><select name="scp3">
<option value="density">density</option>
<option value="histogram">histogram</option>
<option value="boxplot">boxplot</option>
<option value="qqplot">qqplot</option>
<option value="none">none</option></select></td>
</tr></table>
<br><center><div id="scate" class="shiny-plot-output" style="width:
80%; height: 400px"></div></center>
<br><b>2. Koefisien Determinasi ( $R^2$ )</b>
<br><p align="justify">Koefisien determinasi dalam regresi linier
sering diartikan sebagai seberapa besar kemampuan variabel prediktor
dalam menjelaskan varians dari variabel responnya. Koefisien
determinasi juga dapat menjadi indikator kebaikan suatu model. Dalam
hal ini, nilai koefisien determinasi didapat dengan memodelkan
variabel respon (Y) dengan variabel prediktor (X) menggunakan model
linier. Sehingga, nilai  $R^2$  yang lebih besar menunjukkan variabel
prediktor yang lebih cocok jika digunakan pendekatan linier yang
kemudian dimodelkan secara parametrik, dan begitu pula
sebaliknya.</p>
<b>Pilih Variabel Prediktor (X)</b>
<table bgcolor="lightgreen" width="100%">
<tr><td><div id="var.rx" class="shiny-html-output"></div>
</tr></td></table>
<br><b>Nilai  $R^2$  sebesar:</b>
<font color="blue"><pre id="cek.r" class="shiny-text-
output"></pre></font>

<h4><a name="cpd"><b><FONT COLOR="#000000">Cara Pemilihan
Distribusi</FONT></b></a></h4>
<p align="justify">Distribusi yang digunakan dalam model aditif
tergeneralisasi adalah distribusi keluarga eksponensial. Terdapat 4
(empat) distribusi yang disediakan dalam program ini, yaitu
distribusi normal, gamma, poisson, dan binomial. Untuk data kontinu,
dapat menggunakan distribusi normal atau gamma. Sedangkan untuk data
diskrit (cacahan), dapat menggunakan distribusi poisson atau
binomial. Pemilihan distribusi yang cocok, dapat dilihat dari nilai
AIC yang minimum.</p>
<b>Pilih Distribusi</b>
<table bgcolor="lightgreen" width="100%">

```

```

<tr><td><select name="dist.yy">
<option value="NO">Normal</option>
<option value="GA">Gamma</option>
<option value="PO">Poisson</option>
<option value="BB">Binomial</option></select></td></table>
<br><b>Nilai AIC sebesar:</b>
<font color ="blue">
<pre id="pr.hdst" class="shiny-text-output"></pre></font>
<br>Visualisasi Distribusi:
<center><div id="visualdis" class="shiny-plot-output" style="width:
80%; height: 400px"></div>
Gambar Histogram Distribusi</center>

<h3><a name="pmgam"><b><FONT COLOR="#000000"> Pencocokan Model
Menggunakan GAM dengan Penghalus <em>Spline</em></FONT></b></a></h3>
<b>Evaluasi variabel respon (Y) dan variabel prediktor (X) yang
dipilih:</b>
<br><table bgcolor="lightgreen" width="100%">
<tr><td><table bgcolor="lightgreen" width="65%">
<tr><td><center>Variabel Y
<div id="varselectyy" class="shiny-html-output"></div></td></center>
<td><center>Variabel X (komponen parametrik)
<div id="pr.vax" class="shiny-html-output"></div></td></center>
<td><center>Variabel X (komponen nonparametrik)
<div id="npr.vax" class="shiny-html-output"></div></center>
</tr></td></table></td></tr></table>

<br><b>Tentukan:</b>
<br><table bgcolor="lightgreen" width="100%">
<tr><td><table bgcolor="lightgreen" width="65%">
<tr><td>Distribusi (<em>link</em>):
<select name="dist.y">
<option value="gaussian">Normal</option>
<option value="Gamma">Gamma (invers)</option>
<option value="Gammal">Gamma (log)</option>
<option value="poisson">Poisson</option>
<option value="binomial">Binomial (logit)</option>
<option value="binomialp">Binomial (probit)</option>
</select></td>
<td>Basis <em>Spline</em>:
<select name="bs.s">
<option value="cr">Cubic Regression</option>
<option value="cc">Cyclic Cubic Regression</option>
<option value="cs">Cubic Regression (shrinkage)</option>
<option value="ps">Penalized Spline</option>
<option value="tp">Thin Plate Regression</option>
<option value="ts">Thin Plate Regression (shrinkage)</option>
<option value="none">Tanpa Penghalus</option>
</select></td>
<td>Dimensi basis:
<input type="number" name="db.s" value="4" width="1" />
</tr></td></table>

```

```
</tr></td></table>
```

```
<h4><a name="hpmgam"><b><FONT COLOR="#000000"> Hasil Pencocokan
Model Menggunakan GAM dengan Penghalus
<em>Spline</em></FONT></b></a></h4>
```

```
<b>Nilai AIC model sebesar:</b></br>
```

```
<font color ="blue">
```

```
<pre id="aic.m" class="shiny-text-output"></pre></font>
```

```
<b>Ringkasan Hasil Analisis:</b>
```

```
<font color ="blue">
```

```
<pre id="pm" class="shiny-text-output"></pre></font>
```

```
<h4><a name="pmm"><b><FONT COLOR="#000000"> Detail Pilihan Hasil
Pencocokan Model</FONT></b></a></h4>
```

```
<p align="justify">Selain output di atas, terdapat beberapa output
lain yang dapat dipilih mengenai hasil pencocokan model menggunakan
GAM seperti berikut.</p>
```

```
<b>Pilih Output:</b>
```

```
<div id="varselectgam.out" class="shiny-html-output"></div>
```

```
<font color ="blue">
```

```
<pre id="gmout" class="shiny-text-output"></pre></font>
```

```
<h4><a name="pdk"><b><FONT COLOR="#000000"> Term plot
</FONT></b></a></h4>
```

```
<center><div id="plot.diag" class="shiny-plot-output" style="width:
80%; height: 400px"></div></center>
```

```
<h4><a name="us"><b><FONT COLOR="#000000"> Uji
Signifikansi</FONT></b></a></h4>
```

```
<p align="justify">Uji signifikansi dilakukan untuk mengetahui
variabel-variabel prediktor yang signifikan atau memberikan pengaruh
terhadap variabel respon. Uji signifikansi ini dapat dilakukan
dengan menggunakan nilai <em>p-value</em>. Variabel prediktor
dikatakan signifikan jika nilai <em>p-value</em>  $\leq \alpha$ , dan
begitupun sebaliknya. Dimana  $\alpha$  adalah taraf signifikansi yang
bernilai 0,1, 0,05, atau 0,1. Nilai <em>p-value</em> ini diperoleh
dan dapat dilihat pada bagian hasil pencocokan model di atas.</p>
```

```
<h2><a name="Kesimpulan"><b><FONT COLOR="#000000"><b>KESIMPULAN</b>
</FONT></b></a></h2>
```

```
<p align="justify">Model aditif tergeneralisasi (GAM) memiliki
asumsi yang lebih fleksibel jika dibandingkan dengan model linier
maupun model linier tergeneralisasi (GLM), yaitu distribusi yang
digunakan merupakan distribusi keluarga eksponensial dan hubungan
antara variabel respon dengan variabel prediktor tidak harus linier.
Fungsi mulus pada GAM diestimasi menggunakan penghalus diagram
pencar, salah satu penghalus yang sering digunakan para peneliti
karena kelebihan yang dimilikinya adalah penghalus <em>spline</em>.
Dalam GAM ini, model terbaik dapat diketahui melalui nilai AIC yang
minimum. Selain itu, model dengan parameter penghalus optimal dapat
dilihat melalui nilai GCV yang minimum.</p>
```

```
<h2><a          name="dafpus"><b><FONT          COLOR="#000000"><b>DAFTAR
PUSTAKA</b></FONT></b></a></h2>
<p align="justify"><ol>
<li><p align="justify">Eubank, W. 1999. <em>Nonparametric Regression
and Spline Smoothing</em>. New York: Marcel Dekker.</p></li>
<li><p align="justify">Hastie, T. J., dan Tibshirani, R. J. 1990.
<em>Generalized Additive Models</em>. London: Chapman &
Hall.</p></li>
<li><p align="justify">Statsoft. 2016. <em>Generalized Additive
Models</em>. http://www.statsoft.com/textbook/generalized-additive-
models [23 Maret 2016].</p></li>
<li><p align="justify">Tirta, I. M. 2009. <em>Analisis Regresi
dengan R</em>. Jember: Universitas Jember.</p></li>
<li><p align="justify">Wood, S. N. 2006. <em>Generalized Additive
Models: An Introduction with R</em>. London: Chapman & Hall/CRC
Press.</p></li></ol>
```


C. Skrip Program server.r

```

library(shiny)
library(datasets)
library(car)
library(gamlss)
library(mgcv)
data(cars)
data(rock)
data(aids)
data(trees)

# Define server logic for random distribution application
shinyServer(function(input, output, session){
header<-reactive({switch(input$header,
  "TRUE"=TRUE,
  "FALSE"=FALSE)})

Dataset1<- reactive({if(input$pilih.dat=="impor"){
  MyData <- input$dataset
  if (is.null(MyData))
    return(NULL)
  return(data.frame(read.csv(MyData$datapath)))}
else{switch(input$pilih.dat,
  "aids"=aids,
  "rock"=rock,
  "trees"=trees,
  "mtcars"=mtcars)})
})

output$ringkasan <- renderPrint({
  if(input$jdt=="sum"){
    summary(dataset1())}
  else{
    print(dataset1())}
})

output$vareselecty <- renderUI({
  if (identical(dataset1(), '') || identical(dataset2(),
    data.frame()))
    return(NULL)
  # Variable selection:
  selectInput("var.yd", "", names(dataset1()),
    names(dataset1()), selectize=FALSE, multiple =FALSE)})

output$pr.hd<- renderPrint({
  attach(dataset1())
  a1<-do.call("histDist",args=list(y=dataset1()[,input$var.yd],
  family=input$dist.yy,main=paste("Variabel Respon",input$var.yd,
  "dengan Distribusi", input$dist.yy)))
  do.call('AIC', args = list(a1))})

```

```

output$visualdis <- renderPlot({
  attach(dataset1())
  do.call("histDist",args=list(y=dataset1()[,input$var.yd],
  family=input$dist.yy,main=paste("Variabel Respon", input$var.yd,
  "dengan Distribusi", input$dist.yy))))

output$vartable <- renderTable({
  if (identical(dataset1(), '') || identical(dataset1(),
  data.frame()))
    return(NULL)
  # Variable selection:
  selectInput("var.xm","",names(dataset1()), names(dataset1()),
  selectize=FALSE, multiple =TRUE))

output$scate<-renderPlot({
  formula1<-formula(paste("~",paste(input$var.xm, collapse= "+")))
  a2<-do.call("scatterplotMatrix",args=list(formula1,diag=
  input$scp3,smooth=TRUE,smoother=gamLine,main=paste("Scatterplo
  t"))))

output$var.rx <- renderUI({
  if (identical(dataset1(), '') || identical(dataset1(),
  data.frame()))
    return(NULL)
  # Variable selection:
  selectInput("var.rrxm","",names(dataset1()), names(dataset1()),
  selectize=FALSE, multiple =FALSE))

output$cek.r <- renderPrint({
  formula2<-formula(paste(input$var.yd,"~",paste(input$var.rrxm))
  cr2<-lm(formula2,data=dataset1())
  summary(cr2)$r.squared})

output$vartableyy <- renderTable({
  if (identical(dataset1(), '') || identical(dataset1(),
  data.frame()))
    return(NULL)
  # Variable selection:
  selectInput("var.ym","",names(dataset1()),names(dataset1()),
  selectize=FALSE, multiple =FALSE))

output$pr.vax <- renderUI({
  if (identical(dataset1(), '') || identical(dataset1(),
  data.frame()))
    return(NULL)
  # Variable selection:
  selectInput("semi.x","",names(dataset1()),names(dataset1()),
  selectize=FALSE, multiple =TRUE))

output$npr.vax <- renderUI({

```

```

if (identical(dataset1(), '') || identical(dataset1(),
      data.frame()))
  return(NULL)
# Variable selection:
selectInput("semi.sx", "", names(dataset1()), names(dataset1()),
selectize=FALSE, multiple=TRUE)})

output$pm<-renderPrint({
  if(input$bs.s=="cr"){
    formula3<-formula(paste(input$var.ym,"~",paste(input$semi.x,
      collapse="+"), "+", paste("s(", input$semi.sx, "", "bs='cr'", "", "k
      =", input$db.s, ") ", collapse="+")))
  }
  else if(input$bs.s=="cc"){
    formula3<-formula(paste(input$var.ym,"~",paste(input$semi.x,
      collapse="+"), "+", paste("s(", input$semi.sx, "", "bs='cc'", "", "k
      =", input$db.s, ") ", collapse="+")))
  }
  else if(input$bs.s=="cs"){
    formula3<-formula(paste(input$var.ym,"~",paste(input$semi.x,
      collapse="+"), "+", paste("s(", input$semi.sx, "", "bs='cs'", "", "k
      =", input$db.s, ") ", collapse="+")))
  }
  else if(input$bs.s=="ps"){
    formula3<-formula(paste(input$var.ym,"~",paste(input$semi.x,
      collapse="+"), "+", paste("s(", input$semi.sx, "", "bs='ps'", "", "k
      =", input$db.s, ") ", collapse="+")))
  }
  else if(input$bs.s=="tp"){
    formula3<-formula(paste(input$var.ym,"~",paste(input$semi.x,
      collapse="+"), "+", paste("s(", input$semi.sx, "", "bs='tp'", "", "k
      =", input$db.s, ") ", collapse="+")))
  }
  else if(input$bs.s=="ts"){
    formula3<-formula(paste(input$var.ym,"~",paste(input$semi.x,
      collapse="+"), "+", paste("s(", input$semi.sx, "", "bs='ts'", "", "k
      =", input$db.s, ") ", collapse="+")))
  }
  else{
    formula3<-formula(paste(input$var.ym,"~",paste(input$semi.x,
      collapse="+")))
  }
  if(input$dist.y=="Gammal"){
    faml=quote(Gamma(log))
  }
  else if(input$dist.y=="binomialp"){
    faml=quote(binomial(probit))
  }
  else{
    faml=quote(input$dist.y)
  }
  aa<-do.call("gam", args=list(formula=formula3, family=faml))
  summary(aa)
})

output$vselectgam.out <- renderUI({
  gam1<-gam(perm~peri+s(shape), data=rock)
  if (identical(gam1, '') || identical(gam1,data.frame()))
    return(NULL)
# Variable selection:
selectInput("gam.out", "", names(gam1), names(gam1), selectize=FALSE,
multiple =FALSE)})

```

```

output$gmout<-renderPrint({
  if(input$bs.s=="cr"){
    formula3<-formula(paste(input$var.ym,"~",paste(input$semi.x,
      collapse="+"),"+",paste("s(",input$semi.sx,",",",", "bs='cr'",",",",", "k
      =",input$db.s,")",",", collapse="+")))
  }
  else if(input$bs.s=="cc"){
    formula3<-formula(paste(input$var.ym,"~",paste(input$semi.x,
      collapse="+"),"+",paste("s(",input$semi.sx,",",",", "bs='cc'",",",",", "k
      =",input$db.s,")",",", collapse="+")))
  }
  else if(input$bs.s=="cs"){
    formula3<-formula(paste(input$var.ym,"~",paste(input$semi.x,
      collapse="+"),"+",paste("s(",input$semi.sx,",",",", "bs='cs'",",",",", "k
      =",input$db.s,")",",", collapse="+")))
  }
  else if(input$bs.s=="ps"){
    formula3<-formula(paste(input$var.ym,"~",paste(input$semi.x,
      collapse="+"),"+",paste("s(",input$semi.sx,",",",", "bs='ps'",",",",", "k
      =",input$db.s,")",",", collapse="+")))
  }
  else if(input$bs.s=="tp"){
    formula3<-formula(paste(input$var.ym,"~",paste(input$semi.x,
      collapse="+"),"+",paste("s(",input$semi.sx,",",",", "bs='tp'",",",",", "k
      =",input$db.s,")",",", collapse="+")))
  }
  else if(input$bs.s=="ts"){
    formula3<-formula(paste(input$var.ym,"~",paste(input$semi.x,
      collapse="+"),"+",paste("s(",input$semi.sx,",",",", "bs='ts'",",",",", "k
      =",input$db.s,")",",", collapse="+")))
  }
  else{
    formula3<-formula(paste(input$var.ym,"~",paste(input$semi.x,
      collapse="+")))
  }
  if(input$dist.y=="Gammal"){
    faml=quote(Gamma(log))
  }
  else if(input$dist.y=="binomialp"){
    faml=quote(binomial(probit))
  }
  else{
    faml=quote(input$dist.y)
  }
  aa<-do.call("gam",args=list(formula=formula3,family=faml))
  aa[input$gam.out])
}

output$aic.m<- renderPrint({
  if(input$bs.s=="cr"){
    formula3<-formula(paste(input$var.ym,"~",paste(input$semi.x,
      collapse="+"),"+",paste("s(",input$semi.sx,",",",", "bs='cr'",",",",", "k
      =",input$db.s,")",",", collapse="+")))
  }
  else if(input$bs.s=="cc"){
    formula3<-formula(paste(input$var.ym,"~",paste(input$semi.x,
      collapse="+"),"+",paste("s(",input$semi.sx,",",",", "bs='cc'",",",",", "k
      =",input$db.s,")",",", collapse="+")))
  }
  else if(input$bs.s=="cs"){
    formula3<-formula(paste(input$var.ym,"~",paste(input$semi.x,
      collapse="+"),"+",paste("s(",input$semi.sx,",",",", "bs='cs'",",",",", "k
      =",input$db.s,")",",", collapse="+")))
  }
  else if(input$bs.s=="ps"){

```

```
formula3<-formula(paste(input$var.ym,"~",paste(input$semi.x,
collapse="+"),"+",paste("s(",input$semi.sx,",",",", "bs='ps'",",",",", "k
=",input$db.s,")",collapse="+")))
else if(input$bs.s=="tp"){
formula3<-formula(paste(input$var.ym,"~",paste(input$semi.x,
collapse="+"),"+",paste("s(",input$semi.sx,",",",", "bs='tp'",",",",", "k
=",input$db.s,")",collapse="+")))
else if(input$bs.s=="ts"){
formula3<-formula(paste(input$var.ym,"~",paste(input$semi.x,
collapse="+"),"+",paste("s(",input$semi.sx,",",",", "bs='ts'",",",",", "k
=",input$db.s,")",collapse="+")))
else{
formula3<-formula(paste(input$var.ym,"~",paste(input$semi.x,
collapse="+")))
if(input$dist.y=="Gamma"){
faml=quote(Gamma(log))
}
else if(input$dist.y=="binomialp"){
faml=quote(binomial(probit))
}
else{
faml=quote(input$dist.y)
}
aa<-do.call("gam",args=list(formula=formula3,family=faml))
do.call('AIC', args = list(aa)))

output$plot.diag<- renderPlot({
if(input$bs.s=="cr"){
formula3<-formula(paste(input$var.ym,"~",paste(input$semi.x,
collapse="+"),"+",paste("s(",input$semi.sx,",",",", "bs='cr'",",",",", "k
=",input$db.s,")",collapse="+")))
}
else if(input$bs.s=="cc"){
formula3<-formula(paste(input$var.ym,"~",paste(input$semi.x,
collapse="+"),"+",paste("s(",input$semi.sx,",",",", "bs='cc'",",",",", "k
=",input$db.s,")",collapse="+")))
}
else if(input$bs.s=="cs"){
formula3<-formula(paste(input$var.ym,"~",paste(input$semi.x,
collapse="+"),"+",paste("s(",input$semi.sx,",",",", "bs='cs'",",",",", "k
=",input$db.s,")",collapse="+")))
}
else if(input$bs.s=="ps"){
formula3<-formula(paste(input$var.ym,"~",paste(input$semi.x,
collapse="+"),"+",paste("s(",input$semi.sx,",",",", "bs='ps'",",",",", "k
=",input$db.s,")",collapse="+")))
}
else if(input$bs.s=="tp"){
formula3<-formula(paste(input$var.ym,"~",paste(input$semi.x,
collapse="+"),"+",paste("s(",input$semi.sx,",",",", "bs='tp'",",",",", "k
=",input$db.s,")",collapse="+")))
}
else if(input$bs.s=="ts"){
formula3<-formula(paste(input$var.ym,"~",paste(input$semi.x,
collapse="+"),"+",paste("s(",input$semi.sx,",",",", "bs='ts'",",",",", "k
=",input$db.s,")",collapse="+")))
}
else{
formula3<-formula(paste(input$var.ym,"~",paste(input$semi.x,
collapse="+")))
}
if(input$dist.y=="Gamma"){
```



```
faml=quote(Gamma(log))}
else if(input$dist.y=="binomialp"){
  faml=quote(binomial(probit))}
else{
  faml=quote(input$dist.y)}
aa<-do.call("gam",args=list(formula=formula3,family=faml))
a2<-do.call("plot",args=list(aa,pages=1,residuals=TRUE,
  seWithMean=TRUE,pers=TRUE,all.terms=TRUE,main="Term
  plot"))})
})
```

