

SKRIPSI

**PERBANDINGAN OPTIMALISASI MODEL REGRESI RIDGE
MENGUNAKAN METODE ALGORITMA SCHALL DAN BIC**

(Studi Kasus: Model Regresi Ridge untuk Produk Domestik Regional Bruto
(PDRB) Provinsi Jawa Tengah)

***COMPARISON OF RIDGE REGRESSION MODEL OPTIMIZATION USING
SCHALL'S ALGORITHM AND BIC***

*(Case Study: Ridge Regresion Model for Produk Domestik Regional Bruto
(PDRB) Provinsi Jawa Tengah)*



DIWANTI PANCA SATITI

NIM.17106010049

PROGRAM STUDI MATEMATIKA

FAKULTAS SAINS DAN TEKNOLOGI

UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA

YOGYAKARTA

2021

SKRIPSI
PERBANDINGAN OPTIMALISASI MODEL REGRESI RIDGE
MENGGUNAKAN METODE ALGORITMA SCHALL DAN BIC
(Studi Kasus: Model Regresi Ridge untuk Produk Domestik Regional Bruto
(PDRB) Provinsi Jawa Tengah)

COMPARISON OF RIDGE REGRESSION MODEL OPTIMIZATION USING
SCHALL'S ALGORITHM AND BIC

(Case Study: Ridge Regresion Model for Produk Domestik Regional Bruto
(PDRB) Provinsi Jawa Tengah)

Diajukan untuk memenuhi salah satu syarat memperoleh derajat
Sarjana Matematika



DIWANTI PANCA SATITI

NIM.17106010049

PROGRAM STUDI MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA
YOGYAKARTA
2021



SURAT PERSETUJUAN SKRIPSI/TUGAS AKHIR

Hal : Persetujuan Skripsi / Tugas Akhir

Lamp :

Kepada

Yth. Dekan Fakultas Sains dan Teknologi

UIN Sunan Kalijaga Yogyakarta

di Yogyakarta

Assalamu'alaikum wr. wb.

Setelah membaca, meneliti, memberikan petunjuk dan mengoreksi serta mengadakan perbaikan seperlunya, maka kami selaku pembimbing berpendapat bahwa skripsi Saudara:

Nama : Diwanti Panca Satiti

NIM : 17106010049

Judul Skripsi : Optimalisasi Model Regresi Ridge Menggunakan Metode Algoritma Schall dan BIC
sudah dapat diajukan kembali kepada Program Studi Matematika Fakultas Sains dan Teknologi UIN
Sunan Kalijaga Yogyakarta sebagai salah satu syarat untuk memperoleh gelar
Sarjana Strata Satu dalam Program Studi Matematika.

Dengan ini kami mengharap agar skripsi/tugas akhir Saudara tersebut di atas dapat segera dimunaqsyahkan. Atas perhatiannya kami ucapkan terima kasih.

Wassalamu'alaikum wr. wb.

Yogyakarta, 09 Agustus 2021

Sri Utami Zuliana., S.Si., M.Sc., Ph.D.

NIP. 19741003 200003 2 002



KEMENTERIAN AGAMA
UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA
FAKULTAS SAINS DAN TEKNOLOGI
Jl. Marsda Adisucipto Telp. (0274) 540971 Fax. (0274) 519739 Yogyakarta 55281

PENGESAHAN TUGAS AKHIR

Nomor : B-1581/Un.02/DST/PP.00.9/08/2021

Tugas Akhir dengan judul : PERBANDINGAN OPTIMALISASI MODEL REGRESI RIDGE MENGGUNAKAN METODE ALGORITMA SCHALL DAN BIC (Studi Kasus: Model Regresi Ridge untuk Produk Domestik Regional Bruto (PDRB) Provinsi Jawa Tengah)

yang dipersiapkan dan disusun oleh:

Nama : DIWANTI PANCA SATITI
Nomor Induk Mahasiswa : 17106010049
Telah diujikan pada : Jumat, 13 Agustus 2021
Nilai ujian Tugas Akhir : A-

dinyatakan telah diterima oleh Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta

TIM UJIAN TUGAS AKHIR



Ketua Sidang
Sri Utami Zuliana, S.Si., M.Sc., Ph.D.
SIGNED

Valid ID: 612398090d748



Penguji I
Dr. Muhammad Wakhid Musthofa, S.Si.,
M.Si.
SIGNED

Valid ID: 611cb80c90c98



Penguji II
Sri Istiyarti Uswatun Chasanah, M.Si.
SIGNED

Valid ID: 612428c6e717d



Yogyakarta, 13 Agustus 2021
UIN Sunan Kalijaga
Dekan Fakultas Sains dan Teknologi
Dr. Dra. Hj. Khurul Wardati, M.Si.
SIGNED

Valid ID: 6124b115533eb

SURAT PERNYATAAN KEASLIAN

Yang bertanda tangan di bawah ini:

Nama : Diwanti Panca Satiti

NIM : 17106010049

Program Studi : Matematika

Fakultas : Sains dan Teknologi

Dengan ini menyatakan bahwa isi skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar sarjana di suatu Perguruan Tinggi dan sesungguhnya skripsi ini merupakan hasil pekerjaan penulis sendiri sepanjang pengetahuan penulis, bukan duplikasi atau saduran dari karya orang lain kecuali bagian tertentu yang penulis ambil sebagai bahan acuan. Apabila terbukti pernyataan ini tidak benar, sepenuhnya menjadi tanggung jawab penulis.

Yogyakarta, Agustus 2021

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA



Diwanti Panca Satiti

MOTTO

لَا يُكَلِّفُ اللَّهُ نَفْسًا إِلَّا وُسْعَهَا لَهَا مَا كَسَبَتْ وَعَلَيْهَا مَا اكْتَسَبَتْ رَبَّنَا لَا تُؤَاخِذْنَا إِنْ نَسِينَا أَوْ أَخْطَأْنَا
رَبَّنَا وَلَا تَحْمِلْ عَلَيْنَا إَصْرًا كَمَا حَمَلْتَهُ عَلَى الَّذِينَ مِنْ قَبْلِنَا رَبَّنَا وَلَا تُحَمِّلْنَا مَا لَا طَاقَةَ لَنَا بِهِ
وَاعْفُ عَنَّا وَاعْفُ لَنَا وَارْحَمْنَا أَنْتَ مَوْلَانَا فَانصُرْنَا عَلَى الْقَوْمِ الْكَافِرِينَ

Allah tidak membebani seseorang melainkan sesuai dengan kesanggupannya. Dia mendapat (pahala) dari (kebajikan) yang dikerjakannya dan dia mendapat (siksa) dari (kejahatan) yang diperbuatnya. (Mereka berdoa), "Ya Tuhan kami, janganlah Engkau hukum kami jika kami lupa atau kami melakukan kesalahan. Ya Tuhan kami, janganlah Engkau bebani kami dengan beban yang berat sebagaimana Engkau bebani orang-orang sebelum kami. Ya Tuhan kami, janganlah Engkau pikulkan kepada kami apa yang tidak sanggup kami memikulnya. Maafkanlah kami, ampunilah kami, dan rahmatilah kami. Engkaulah pelindung kami, maka tolonglah kami menghadapi orang-orang kafir."

(QS. Al-Baqarah 2: Ayat 286)

"Jika kamu tidak sanggup menahan lelahnya belajar maka kamu harus sanggup menahan perihnya kebodohan."

~Imam Syafi'i~

Apa yang melewatkanmu tidak akan pernah menjadi takdirku dan apa yang ditakdirkan untukku tidak akan pernah melewatkanmu.

~Umar bin Khattab~

Tidak sempurna bukan berarti diam pada kesalahan yang sama.

HALAMAN PERSEMBAHAN

Alhamdulillah Rabbil 'aalamiin.

Karya ini saya persembahkan untuk:

Keluarga yang selalu memberikan do'a dan dukungan
fasilitas

Seluruh orang-orang baik yang secara sengaja maupun
tidak sengaja telah memberikan suasana positif dalam
kehidupan saya

Teman-teman satu bimbingan yang tiada henti memberikan
semangat, mendukung dan memberikan kritik maupun
saran.

Julita Sari atas bantuan laptop dan R studionya

Seluruh sahabat-sahabat saya

Keluarga Besar Program Studi Matematika Fakultas Sains
dan Teknologi Universitas Islam Negeri Sunan Kalijaga
Yogyakarta



KATA PENGANTAR

Puji dan syukur saya panjatkan kepada Allah SWT karena telah memberikan rohmat dan hidayatnya kepada saya sehingga saya dapat menyelesaikan skripsi ini guna memenuhi persyaratan memperoleh derajat sarjana S1 Program Studi Matematika (S.Mat) pada Fakultas Sains dan Teknologi Universitas Islam Negeri Sunan Kalijaga.

Pada halaman kata pengantar ini, saya juga ingin mengucapkan banyak terima kasih kepada pihak-pihak yang telah memberikan dukungan, semangat, doa, dan motivasi kepada saya sehingga saya dapat menyelesaikan skripsi ini.

1. Dr. Khurul Wardati, selaku Dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Sunan Kalijaga.
2. Muchammad Abrori, S.Si.,M.Kom, selaku ketua Program Studi Prodi Matematika Universitas Islam Negeri Sunan Kalijaga.
3. Muhamad Zaki Riyanto, S.Si., M.Sc, selaku dosen pembimbing akademik.
4. Sri Utami Zuliana., S.Si.,M.Sc., Ph.D, selaku dosen pembimbing skripsi.
Terima kasih atas kesabarannya dalam membimbing saya, terima kasih pula atas waktu, arahan, dan masukan yang diberikan. Ilmu yang diberikan sangat bermanfaat untuk pengerjaan skripsi ini.
5. Bapak dan Ibu dosen Prodi Matematika Universitas Islam Negeri Sunan Kalijaga yang telah memberikan banyak ilmu dan berbagi pengalaman hidup yang InsyaAllah akan bermanfaat bagi saya.

6. Seluruh staff Prodi Matematika Universitas Islam Negeri Sunan Kalijaga dan jajarannya yang telah mempermudah saya dalam pengurusan syarat-syarat seminar proposal dan sidang saya.
7. Untuk segenap keluarga besar saya. Terima kasih karena telah memberikan dukungan kepada saya selama kuliah, untuk biaya-biaya yang telah dikeluarkan hingga saya dapat melangkah sejauh ini dalam menempuh pendidikan.
8. Untuk teman-teman satu bimbingan saya Anisa Naswa Listiani, Istinganatun, Linaksanan dan Royhana Devi yang sering memberikan semangat, masukan, kritik, saran, dan selalu mengajak saya untuk mengerjakan skripsi bersama agar skripsi ini dapat segera selesai.
9. Untuk sahabat-sahabat saya Royhana Devi, Hannifa Irmajihan Nabila, Anatansyah Ayomi Anandari, Riyana Yuni Sulistiyowati, Julita Sari, Novia Amilatus Sholihah, dan teman-teman Matematika UIN Sunan Kalijaga. Terima kasih atas cerita-cerita berkesan yang telah kalian berikan selama masa perkuliahan saya.
10. Untuk teman – teman KKN Sambas saya Hadi Wiryawan, Lia Nurhayati, Annisa Robbiatun Tammimah, Yusika Intan Insiwi, Delina Restiani, Amien Fajar Khuzaeni, Indri Nur Anggraini, Sukma Wahyuni, dan Naspadina. Terima kasih karena kalian telah begitu ambisius untuk mengejar mimpi-mimpi kalian sehingga saya bisa ikut termotivasi untuk segera menyusul lulus dari UIN Sunan Kalijaga seperti kalian yang sudah lulus terlebih dahulu..

11. Untuk Ezra Afrizardi Yusufa. Terima kasih karena telah mengajarkan saya dalam memahami R.

12. Untuk teman-teman pendakian saya yang namanya tidak bisa saya sebut satu persatu. Terima kasih karena sering mengajak saya keluar untuk menyegarkan pikiran saya, untuk sejenak *mengcharge* tenaga sebelum kembali mengerjakan skripsi ini.

Semoga skripsi ini dapat bermanfaat bagi pembaca. Penulis memohon maaf apabila terdapat kesalahan dalam skripsi ini. Penulis akan selalu menerima kritik serta saran yang pembaca berikan.

Yogyakarta, 30 Juli 2021

Diwanti Panca Satiti

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

DAFTAR ISI

JUDUL	i
SURAT PERSETUJUAN SKRIPSI/TUGAS AKHIR	iii
HALAMAN PENGESAHAN.....	iv
HALAMAN PERSEMBAHAN	v
MOTTO	vi
DAFTAR ISI.....	xi
DAFTAR TABEL.....	xiv
DAFTAR LAMPIRAN.....	xv
DAFTAR GAMBAR	xvi
DAFTAR SIMBOL.....	xvii
INTISARI.....	xviii
ABSTRACT.....	xix
BAB I	1
PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	3
1.3. Tujuan Penelitian.....	3
1.4. Manfaat Penelitian.....	4
1.5. Sistematika Penulisan.....	4
BAB II.....	7
TINJAUAN PUSTAKA	7
BAB III	9

RANCANGAN PENELITIAN	9
3.1. Landasan Teori	9
3.1.1. Analisis Regresi	9
3.1.2. Analisis Regresi Linier Berganda	10
3.1.3. Multikolinearitas	12
3.1.4. Regresi Ridge	16
3.1.5. <i>Bayesian Information Criterion (BIC)</i>	17
3.1.6. Algoritma Schall	18
3.2. Hipotesis	19
3.3. Rancangan Penelitian	20
3.3.1. Jenis Penelitian	20
3.3.2. Tahapan penelitian	20
BAB IV	22
METODE PENELITIAN	22
4.1. Pendekatan Penelitian	22
4.2. Data	22
4.3. Variabel dan Definisi Variabel	23
4.4. Alat dan Cara Penelitian	23
BAB V	25
HASIL DAN PEMBAHASAN	25
5.1. Analisis Deskriptif	25
5.1.1. Data Laju Pertumbuhan PDRB	25
5.1.2. Rata-rata	27
5.1.3. Variansi	27
5.2. Analisis Regresi	28

5.2.1. Uji F	29
5.2.2. Uji t	30
5.3. Mendeteksi Adanya Multikolinearitas	33
5.4. Model Regresi Ridge Terbaik dengan Menggunakan Algoritma Schall	35
5.5. Model Regresi Ridge Terbaik dengan Menggunakan (BIC).....	37
5.6. Kriteria Pemilihan Model Terbaik	38
BAB VI	40
KESIMPULAN	40
6.1. Kesimpulan.....	40
6.2. Saran.....	41
LAMPIRAN.....	44
CURRICULUM VITAE	69

DAFTAR TABEL

No	Judul	Halaman
3.1	Besar konsumsi di Amerika Serikat	17
3.2	Nilai VIF variabel besar konsumsi di Amerika Serikat	17
4.1	Definisi variabel	23
5.1	Data laju pertumbuhan PDRB Provinsi Jawa Tengah menurut lapangan usaha	25
5.2	Rata-rata data laju pertumbuhan PDRB Provinsi Jawa Tengah menurut lapangan usaha	27
5.3	Variansi data laju pertumbuhan PDRB Provinsi Jawa Tengah menurut lapangan usaha	28
5.4	Nilai VIF laju pertumbuhan PDRB Provinsi Jawa Tengah menurut lapangan usaha	34
5.5	Nilai MSE dan MAPE model menggunakan algoritma Schall dan <i>BIC</i>	38

DAFTAR LAMPIRAN

No	Judul	Halaman
1	Program menentukan nilai VIF variabel independen besar konsumsi di Amerika Serikat	45
2	Output program menentukan nilai VIF variabel independen besar konsumsi di Amerika Serikat	46
3	Program menentukan koefisien regresi	47
4	Output program menentukan koefisien regresi	48
5	Program uji normalitas	50
6	Output program uji normalitas	51
7	Program regresi ridge	52
8	Output program regresi ridge	54
9	Program algoritma Schall dan BIC	56
10	Output program algoritma Schall dan BIC	61
11	Data PDRB Provinsi Jawa Tengah Menurut Lapangan Usaha	68

DAFTAR GAMBAR

No	Judul	Halaman
5.1.	Plot residu data laju pertumbuhan PDRB	31
5.2.	Grafik pengiterasian algoritma Schall	36



DAFTAR SIMBOL

Y	: Variabel terikat atau variabel dependen
X_k	: Variabel bebas atau variabel independen ke-k
β_0	: Intercept
β_p	: Koefisien regresi pada variabel X_p
ε	: Variabel pengganggu/ residual
$\sum X$	: Jumlah semua nilai data
N	: Banyaknya data sampel
σ^2	: Variansi Populasi
S^2	: Variansi deviasi sampel
x	: Data
μ	: Rata-rata data populasi
$\bar{x}$	: Rata-rata data sampel
N	: Jumlah data populasi
P	: Jumlah parameter model ke-m
$\hat{\lambda}$	: Bobot penalty optimal

OPTIMALISASI MODEL REGRESI RIDGE MENGUNAKAN METODE ALGORITMA SCHALL DAN BIC

(Studi Kasus: Model Regresi Ridge untuk Produk Domestik Regional
Bruto (PDRB) Provinsi Jawa Tengah)

Oleh: Diwanti Panca Satiti

INTISARI

Sebuah model regresi dikatakan baik jika telah memenuhi asumsi-asumsi klasik salah satunya yaitu tidak adanya multikolinearitas. Jika ditemukan adanya masalah multikolinearitas terdapat beberapa prosedur yang dapat digunakan untuk mengatasi masalah multikolinearitas tersebut salah satunya dengan menggunakan model regresi ridge.

Dalam penelitian ini, menentukan model terbaik pada regresi ridge dilakukan dengan metode *Bayesian Information Criterion (BIC)* dan algoritma Schall. Langkah menentukan model terbaik dengan menggunakan algoritma Schall diawali dengan memilih sembarang bobot penalti awal untuk mengestimasi koefisien regresi ridge. Sedangkan pada metode *BIC* langkah menentukan model terbaik dihitung dari beberapa bobot penalti dan model terbaik memiliki nilai *BIC* terkecil.

Hasil dari penelitian ini diperoleh metode algoritma Schall lebih efektif untuk mengoptimisasi model. Hal ini dikarenakan nilai *MSE* dan nilai *MAPE* metode algoritma Schall lebih kecil daripada nilai *MSE* dan nilai *MAPE* metode *BIC*.

Kata Kunci : Algoritma Schall, *BIC*, *MAPE*, *MSE*, PDRB, Provinsi Jawa Tengah, Regresi Ridge

OPTIMIZATION OF RIDGE REGRESSION MODEL USING SCHALL'S ALGORITHM AND BIC

(Case Study: Ridge Regresion Model for Produk Domestik Regional
Bruto (PDRB) Provinsi Jawa Tengah)

by: Diwanti Panca Satiti

ABSTRACT

Linear regression modelshave some assumptions.One of them is the absence of multicollinearity. If there is a multicollinearity problem, there are several procedures that can be used to overcome the multicollinearity problem. Ridge regression models are one of models could be applied for a multicollinear dataset.

In this research, selecting the best model for ridge regression will be applied using Bayesian Information Criterion (BIC) and Schall's Algorithm. Schall's algorithm begins with choosing any initial penalty weight to estimate the ridge regression coefficient.While BIC is calculated from some penalty weights and the best model has the smallest BIC.

The results of this study obtained that the Schall's algorithm is more effective for optimizing the model than BIC. MSE and MAPE of the Schall's algorithm smaller than MSE and MAPE of the BIC.

Keywords: *Schall's Algorithm, BIC, MAPE, MSE, GRDP*, Central Java Province, Ridge Regression

BAB I

PENDAHULUAN

Pada bab pendahuluan ini akan dijabarkan tentang latar belakang dilakukan penelitian. Selain itu juga dimuat tujuan penelitian yang menjelaskan secara rinci tentang apa yang hendak tercapai dengan adanya penelitian ini, manfaat penelitian baik bagi penulis maupun pembaca, dan sistematika penulisan.

1.1. Latar Belakang

Statistika merupakan disiplin ilmu yang mempelajari sekumpulan konsep dan metode pengumpulan, penyajian, analisis, dan interpretasi data sampai dengan pengambilan keputusan pada situasi dimana terdapat ketidakpastian. Dalam statistika, ada yang dinamakan analisis regresi yang merupakan alat statistik untuk membuat perkiraan (prediksi) yang dapat dipercaya untuk suatu variabel jika nilai variabel lain yang berhubungan dengannya diketahui. Sebuah model regresi dikatakan baik jika telah memenuhi asumsi-asumsi klasik yaitu tidak adanya autokorelasi, heteroskedastisitas dan multikolinearitas.

Multikolinearitas terjadi ketika menentukan model regresi populasi ada kemungkinan bahwa dalam sampel tertentu, beberapa atau semua variabel sangat kolinear (mempunyai hubungan linear sempurna atau hampir sempurna). Ada beberapa prosedur yang dapat digunakan untuk mengatasi masalah multikolinearitas. Apabila terdapat unsur multikolinearitas dalam data dengan derajat yang bisa ditolerir, kita dapat membuang atau menyeleksi variabel selama hal tersebut tidak mengubah teori yang ada (variabel tidak begitu penting

kehadirannya di dalam model). Namun, tidak semua permasalahan multikolinearitas dapat menggunakan metode tersebut karena tak jarang metode tersebut akan mempengaruhi variabel tak bebas. Apabila metode tersebut dapat mempengaruhi variabel tak bebas, ada beberapa prosedur yang dapat dipilih diantaranya memperbesar ukuran sampel, memasukkan persamaan tambahan ke dalam model, penggunaan informasi ekstra, dan dengan analisis regresi ridge.

Salah satu dari cara mengatasi permasalahan multikolinearitas tersebut ialah dengan menggunakan metode regresi ridge. Metode regresi ridge adalah salah satu metode regresi terpenalti, metode regresi ridge menghasilkan taksiran koefisien regresi dengan varian yang lebih kecil namun taksiran koefisien regresinya bersifat bias. Terdapat beberapa metode yang dapat digunakan untuk menentukan model terbaik pada regresi ridge hingga saat ini seperti *Akaike Information Criterion (AIC)* (Akaike, 1974), *Bayesian Information Criterion (BIC)* (Schwarz, 1978), *Generalized Cross Validation (GCV)* (Golub et al., 1979), dan algoritma Schall (Zuliana S.U, dan Perperoglou A., 2016).

Pada penelitian ini digunakan metode *BIC* dan algoritma Schall untuk menentukan model terbaik pada regresi ridge. Metode *BIC* dipilih sebab dirasa cukup akurat karena memperhatikan jumlah parameter dalam model. Sedangkan algoritma Schall dipilih sebab algoritma Schall merupakan hasil dari penelitian yang dilakukan setelah penelitian konsep *BIC* pada tahun 1978 yang dikemukakan oleh Schwartz. Berdasar dari latar belakang diatas tersebut maka penulis tertarik untuk melakukan penelitian “Optimalisasi Model Regresi Ridge Menggunakan Metode algoritma Schall dan *BIC*”

1.2. Rumusan Masalah

1. Bagaimana model regresi ridge dari laju pertumbuhan PDRB Provinsi Jawa Tengah?
2. Bagaimana langkah-langkah dalam menentukan pemilihan model terbaik dengan metode algoritma Schall dan *Bayesian Information Criterion (BIC)*?
3. Bagaimana perbandingan keefektifan algoritma Schall dengan *Bayesian Information Criterion (BIC)* dalam pemilihan model terbaik pada studi kasus Produk Domestik Regional Bruto (PDRB) Provinsi Jawa tengah?

1.3. Tujuan Penelitian

Tujuan yang hendak dicapai dalam penelitian ini adalah

1. Mengkaji model regresi ridge.
2. Mengkaji pemilihan model terbaik dengan metode algoritma Schall dan *Bayesian Information Criterion (BIC)*.
3. Mengetahui perbandingan keefektifan algoritma Schall dengan *Bayesian Information Criterion (BIC)* dalam pemilihan model terbaik pada studi kasus Produk Domestik Regional Bruto (PDRB) Provinsi Jawa tengah.

1.4. Manfaat Penelitian

Bagi penulis :

1. Dapat mengkaji ulang metode algoritma Schall dan menerapkannya dalam pemilihan model regresi terbaik.
2. Dapat mengkaji ulang metode *Bayesian Information Criterion* dan menerapkannya dalam pemilihan model regresi terbaik.
3. Dapat membandingkan seberapa efektif algoritma Schall dengan *Bayesian Information Criterion (BIC)* dalam mengoptimisasi regresi ridge.

Bagi Pembaca :

Dapat memberikan tambahan ilmu tentang cara mengatasi multikolinearitas dengan optimasi algoritma Schall dan *Bayesian Information Criterion (BIC)*.

1.5. Sistematika Penulisan

Penelitian ini terdiri dari beberapa BAB dengan garis-garis besar isi BAB akan dijelaskan di bawah dengan tujuan dapat memudahkan pembaca dalam memahami isi skripsi ini.

BAB I PENDAHULUAN

Pada bab pendahuluan dijabarkan tentang latar belakang dilakukan penelitian. Selain itu, juga dimuat tujuan penelitian yang menjelaskan secara rinci tentang apa yang hendak tercapai dengan adanya penelitian ini, manfaat penelitian baik bagi penulis maupun pembaca, dan sistematika penulisan.

BAB II TINJAUAN PUSTAKA

Pada bab tinjauan pustaka dijabarkan mengenai isi atau kesimpulan yang dapat ditarik dari penelitian sebelumnya yang kemudian menjadikan acuan untuk peneliti melakukan penelitian ini.

BAB III RANCANGAN PENELITIAN

Pada bab rancangan penelitian ini dijabarkan tentang landasan teori yang memuat pengertian, rumus, maupun sifat-sifat dari teori yang diperlukan di bab selanjutnya. Adapun teori-teori yang digunakan dalam penelitian seperti analisis regresi, analisis regresi linier berganda, multikolinearitas, regresi ridge, *bayesian information criterion (BIC)*, dan algoritma Schall. Selain itu pada bab ini juga dijelaskan mengenai hipotesis penelitian yang memuat tinjauan pustaka yang sekaligus menjadi jawaban sementara untuk masalah metode mana yang paling optimal untuk menentukan model terbaik regresi ridge. Pada bab rancangan penelitian juga disebutkan tahapan penelitian dengan gambaran *flowchart* guna mempermudah pembaca mengetahui alur penelitian ini dibuat.

BAB IV METODE PENELITIAN

Pada bab metode penelitian ini disajikan mengenai pendekatan penelitian yang digunakan, sumber dan gambaran umum data yang dipakai untuk menyelesaikan masalah penelitian, variabel dan definisi variabel, dan juga alat serta cara penelitian.

BAB V PEMBAHASAN

Pada bab pembahasan ini dijelaskan mengenai model regresi dari data yang digunakan, analisis deskriptif dari data yang digunakan, pendeteksian adanya

gejala multikolinearitas, cara mengatasi gejala multikolinearitas dengan menggunakan regresi ridge, serta langkah-langkah dan kriteria dalam pemilihan model terbaik dengan metode algoritma Schall dan *Bayesian Information Criterion (BIC)*”.

BAB VI KESIMPULAN

Pada bab kesimpulan ini dijabarkan mengenai hasil yang didapat dari penelitian yang sekaligus merupakan jawaban dari tujuan penelitian. Selain itu, pada bab kesimpulan ini juga disebutkan saran atas kekurangan dalam penelitian dengan tujuan penelitian ini dapat dikembangkan lebih lanjut di kemudian hari.

BAB VI

KESIMPULAN

Setelah data laju pertumbuhan PDRB Provinsi Jawa Tengah menurut lapangan usaha dianalisis dan dijabarkan seperti pada BAB V (bab hasil dan pembahasan) dapat ditarik beberapa kesimpulan dan saran sebagai berikut:

6.1. Kesimpulan

Berdasarkan rumusan masalah dan bab V (hasil dan pembahasan) didapatkan kesimpulan sebagai berikut:

1. Metode regresi ridge pada dasarnya diperoleh dengan cara yang sama seperti metode kuadrat terkecil yaitu dengan meminimumkan jumlah kuadrat sisaan. Namun, pada regresi ridge ditambahkan konstanta bias c pada diagonal $X^T X$. Nilai c untuk koefisien regresi ridge ini antara 0 sampai dengan 1 sehingga koefisien berkurang dan mendekati nol.
2. Terdapat beberapa cara untuk menentukan pemilihan model regresi terbaik diantaranya ialah menggunakan metode *Bayesian Information Criterion (BIC)* dan Algoritma Schall. Langkah menentukan model terbaik dengan menggunakan algoritma Schall diawali dengan memilih sembarang bobot penalty awal untuk mengestimasi koefisien regresi ridge hingga mendapatkan bobot penalti $\hat{\lambda}$ yang optimal. Sedangkan pada metode *BIC*, *BIC* dihitung dengan beberapa bobot penalti kemudian model terbaik adalah model dengan *BIC* terkecil.

3. Algoritma Schall lebih efektif daripada *Bayesian Information Criterion (BIC)* dalam pemilihan model terbaik untuk model regresi ridge PDRB. Model yang dihasilkan dari metode algoritma Schall memiliki nilai *MSE* 146.4836 dan nilai *MAPE* 2.84124 dimana semua nilai ini besarnya lebih kecil dibanding nilai *MSE* dan nilai *MAPE* dari model yang dihasilkan dengan metode *Bayesian Information Criterion (BIC)* yang berturut-turut nilainya sebesar 193.8899 dan 3.252187. Sehingga model regresi ridge terbaik yang dihasilkan $\hat{y} = 0.0165 \text{ PKP} + 0.6377 \text{ PP} + 5.7764 \text{ IP} - 0.0479 \text{ PLG} + 2.6839 \text{ PA} - 5.1202 \text{ K} - 0.5460 \text{ PBE}$

6.2. Saran

Untuk menentukan model terbaik pada regresi ridge telah terdapat beberapa metode yang diperkenalkan seperti *Akaike Information Criterion (AIC)* (Akaike, 1974), *Bayesian Information Criterion (BIC)* (Schwarz, 1978), *Generalized Cross Validation (GCV)* (Golub et al., 1979), dan algoritma Schall (Zuliana S.U, dan Perperoglou A., 2016). Pada penelitian ini peneliti hanya membandingkan dua metode yaitu *Bayesian Information Criterion (BIC)* dan algoritma Schall, sehingga bagi peneliti berikutnya dapat mengembangkan teori ini dengan membandingkan metode algoritma Schall dengan kedua metode yang lainnya.

DAFTAR PUSTAKA

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control*, 19(6), 716-723.
- Badan Pusat Statistik [BPS]. (2021). Beberapa Indikator Makro Sosial Ekonomi Provinsi Jawa Tengah. Jawa tengah, Badan Pusat Statistik Provinsi Jawa Tengah.
- Draper, N. R. Smith. H.(1981). *Applied regression analysis*, 661-81.
- Golub, G. H., Heath, M., & Wahba, G. (1979). Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics*, 21(2), 215-223.
- Hoerl, A. E. (1962). Application of ridge analysis to regression problems. *Chemical Engineering Progress*. 58, 54-59.[41] Hoerl, AE and Kennard, RW (1970) Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*, 12, 55-67.
- Ifadah, A. (2011). *Analisis Metode Principal Component Analysis (Komponen Utama) Dan Regresi Ridge Dalam Mengatasi Dampak Multikolinearitas Dalam Analisis Regresi Linear Berganda*. Skripsi. Jurusan Matematika Fakultas Matematika Dan Ilmu Pengetahuan Alam Universitas Negeri Semarang, Semarang.
- Kustinah, N. (2012). *Pemilihan Model Regresi Terbaik dengan Bayesian Information Criterion (BIC)*. Skripsi. Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Sebelas Maret, Surakarta.

- Nabila, S.U. (2018). *Perbandingan Metode Regresi Ridge dan Metode Regresi Komponen Utama dalam Menangani Multikolinearitas*. Skripsi. Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung, Bandar Lampung.
- Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics*, 461-464.
- Qudratullah, M. F. (2013). Analisis Regresi Terapan: Teori, Contoh Kasus, dan Aplikasi dengan SPSS. *CV Andi Offset, Yogyakarta*.
- Zuliana, S. U. (2018). Penentuan Model Terbaik Regresi Ridge Dan Terapannya. *Jurnal Ilmiah Matematika dan Pendidikan Matematika*, 10(2), 43-48.
- Zuliana, S. U., & Perperoglou, A. (2016). The weight of penalty optimization for ridge regression. In *Analysis of Large and Complex Data* (pp. 231-239). Springer.



LAMPIRAN

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

**Lampiran 1. Program menentukan nilai VIF variabel independen
besar konsumsi di Amerika Serikat**

```
>dataregresi=read.csv("data1.csv",sep=";")  
>regresi=lm(Y~X1+x2+x3,data=dataregresi)  
>regresi  
>summary(regresi)  
>library(car)  
>vif(regresi)
```



**Lampiran 2. Output program menentukan nilai VIF variabel independen
besar konsumsi di Amerika Serikat**

```

>dataregresi=read.csv("data1.csv",sep=",")
>regresi=lm(Y~X1+x2+x3,data=dataregresi)
>regresi

Call:
lm(formula = Y ~ x1 + x2 + x3, data = dataregresi)

Coefficients:
(Intercept)      x1      x2      x3
  18.7021    0.3803    1.4186    0.5331

>summary(regresi)

Call:
lm(formula = Y ~ x1 + x2 + x3, data = dataregresi)

Residuals:
    Min       1Q   Median       3Q      Max
-13.494  -1.847   1.116   2.541   6.460

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  18.7021    6.8454    2.732  0.0211 *
x1           0.3803    0.3121    1.218  0.2511
x2           1.4186    0.7204    1.969  0.0772 .
x3           0.5331    1.3998    0.381  0.7113
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.06 on 10 degrees of freedom
Multiple R-squared:  0.9187,    Adjusted R-squared:  0.8943
F-statistik: 37.68 on 3 and 10 DF,  p-value: 9.271e-06

>library(car)
>vif(regresi)
      x1      x2      x3
12.296544  9.230073  2.976638

```

Lampiran 3. Program menentukan koefisien regresi

```
dataregresi=read.csv("DataPDRB1.csv",sep=",")  
Dataregresi  
regresi=lm(PDRB~PKP+PP+IP+PLG+PA+K+PBE,data=dataregresi)  
Regresi  
summary(regresi)
```



Lampiran 4. Output program menentukan koefisien regresi

```
> dataregresi=read.csv("DataPDRB1.csv",sep=",")
> dataregresi
```

	PDRB	PKP	PP	IP	PLG	PA	K	PBE
1	5.32	10.06	6.70	3.74	4.05	7.19	5.55	5.40
2	5.18	-3.49	7.77	5.04	5.52	6.10	7.08	8.07
3	5.17	-0.22	2.06	4.25	7.61	6.91	7.44	7.18
4	5.40	0.36	4.56	4.34	3.81	5.88	8.33	3.84
5	5.43	1.32	1.85	4.85	4.11	7.11	6.89	4.84
6	5.54	3.52	5.69	4.39	5.36	5.76	6.34	5.58
7	1.75	19.12	-0.25	-0.10	-3.12	1.39	-3.63	0.47
8	3.45	5.49	2.63	3.51	1.26	3.21	1.24	3.98
9	2.48	4.06	-2.10	1.88	2.55	-0.40	7.89	1.23
10	-2.35	-22.35	1.82	-0.09	5.42	0.69	0.16	-1.62
11	-0.84	11.76	2.63	-2.23	-4.44	-0.45	-7.60	-0.27
12	-5.18	11.76	-6.87	-3.95	-4.45	1.32	-5.70	-9.72
13	4.79	9.91	3.30	-0.99	7.93	0.62	8.16	6.58
14	-1.89	-21.65	-2.55	0.98	1.58	1.28	1.45	1.85
15	1.69	18.33	2.20	0.02	-0.04	0.99	-2.59	1.47

```
> regresi=lm(PDRB~PKP+PP+IP+PLG+PA+K+PBE,data=dataregresi)
> regresi
```

Call:
lm(formula = PDRB ~ PKP + PP + IP + PLG + PA + K + PBE, data = dataregresi)

Coefficients:
(Intercept) PKP PP IP
PLG PA K
-0.111185 0.112733 0.111220 0.383730
0.038221 -0.005358 0.194960
PBE
0.255287

```
> summary(regresi)
```

Call:
lm(formula = PDRB ~ PKP + PP + IP + PLG + PA + K + PBE, data = dataregresi)

Residuals:

Min	1Q	Median	3Q	Max
-0.73233	-0.32556	0.01413	0.25918	0.62768

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.111185	0.257794	-0.431	0.67923
PKP	0.112733	0.019367	5.821	0.00065 ***
PP	0.111220	0.078641	1.414	0.20018
IP	0.383730	0.176058	2.180	0.06568 .
PLG	0.038221	0.124331	0.307	0.76748
PA	-0.005358	0.107905	-0.050	0.96178
K	0.194960	0.087579	2.226	0.06132 .
PBE	0.255287	0.092550	2.758	0.02816 *

 Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5777 on 7 degrees of freedom
 Multiple R-squared: 0.9861, Adjusted R-squared: 0.9723
 F-statistik: 71.18 on 7 and 7 DF, p-value: 5.674e-06

Lampiran 5. Program uji normalitas

```
ujishapiro<-shapiro.test(dataregresi$PDRB)
Ujishapiro
norm<-
aov(PDRB~PKP+PP+IP+PLG+PA+K+PBE,data=dataregresi)
qqnorm(norm$residuals)
qqline(norm$residuals, col="red")
```



Lampiran 6. Output program uji normalitas

```
> ujishapiro<-shapiro.test(dataregresi$PDRB)
> ujishapiro
```

Shapiro-wilk normality test

```
data: dataregresi$PDRB
W = 0.84567, p-value = 0.01509
```

```
> norm<-
aov(PDRB~PKP+PP+IP+PLG+PA+K+PBE,data=dataregresi)
> qqnorm(norm$residuals)
> qqline(norm$residuals, col="red")
```

Lampiran 7. Program regresi ridge

```

df<-read.csv("dataPDRB1.csv", sep = ",")
Df
regresi<-lm(PDRB~PKP+PP+IP+PLG+PA+K+PBE, data = df)
summary(regresi)
library(car)
vif(regresi)

carPDRB<-dat$PDRB
X<-dat[, -1]
ff<-function(x){(x-mean(x))/sd(x)}
Xx<-apply(X,2,ff)
Xx
df<-data.frame(PDRB,Xx)
Df
library(ggplot2)
library(ridge)
library(gamlss)
library(penalized)
library(MASS)

#REGRESI RIDGE
Inridge <- function(Y,X,lambda)
{
  n <- nrow(X)
  iot <- diag(ncol(X))
  dz <- 5
  while(dz>1e-4){
    sol <- solve(t(X)%*%X+lambda*iot)
    #RUMUS REGRESI RIDGE
    beta <- sol%*%t(X)%*%Y
    hat <- X %*% sol%*%t(X)
    #SCAHL ALGORITHM
    lam <- sum(diag(hat))/(t(beta)%*%beta)
    dz <- abs(lambda-lam)
    lambda<- as.numeric(lam)
    cat(lam,"\n")
  }
  res <- Y-X%*%beta
  logl <- .5* ( - n * (log(2 * pi) + 1 - log(n) +log(sum(res^2))))

```

```
out <- list("coef"=round(beta,4),"logL"=logl, "hat"=hat, "lambda"=lambda)
Out
}
n<-29
batas <- ceiling(0.8*n)
d.train <- df[1:batas,]
d.test <- df[batas:n,]

fit.r <- linearRidge(PDRB~.-1,lambda="automatic", data = d.train)
lam.r <- fit.r$lambda[2]
lam.r
b.r <- coef(fit.r)
b.r
L.r <- c(b.r, lam.r)
L.r
```



Lampiran 8. Output program regresi ridge

```
> regresi<-lm(PDRB~PKP+PP+IP+PLG+PA+K+PBE, data = df)
> summary(regresi)
```

Call:

```
lm(formula = PDRB ~ PKP + PP + IP + PLG + PA + K + PBE, data = df)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-0.73233	-0.32556	0.01413	0.25918	0.62768

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.111185	0.257794	-0.431	0.67923
PKP	0.112733	0.019367	5.821	0.00065 ***
PP	0.111220	0.078641	1.414	0.20018
IP	0.383730	0.176058	2.180	0.06568 .
PLG	0.038221	0.124331	0.307	0.76748
PA	-0.005358	0.107905	-0.050	0.96178
K	0.194960	0.087579	2.226	0.06132 .
PBE	0.255287	0.092550	2.758	0.02816 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5777 on 7 degrees of freedom

Multiple R-squared: 0.9861, Adjusted R-squared: 0.9723

F-statistik: 71.18 on 7 and 7 DF, p-value: 5.674e-06

```
> library(car)
```

```
> vif(regresi)
```

	PKP	PP	IP	PLG	PA	K	PBE
	2.321189	3.694182	10.554213	10.421756	4.246587	9.740609	7.134514

```
> Inridge <- function(Y,X,lambd)
```

```
+ {
```

```
+ while(dz>1e-4){
```

```
+ sol <- solve(t(X)%*%X+lambd*iot)
```

```
+ #RUMUS REGRESI RIDGE
```

```

+ beta <- sol%*%t(X)%*%Y
+ hat <- X %*% sol%*%t(X)
+ #SCAHL ALGORITHM
+ lam <- sum(diag(hat))/(t(beta)%*%beta)
+ }
+ res <- Y-X%*%beta
+ }
> n<-29
> batas <- ceiling(0.8*n)
> d.train <- df[1:batas,]
> d.test <- df[batas:n,]
>
> fit.r <- linearRidge(PDRB~.-1,lambda="automatic", data = d.train)
> lam.r <- fit.r$lambda[2]
> lam.r
[1] 0.03558539
> b.r <- coef(fit.r)
> b.r
      PKP      PP      IP      PLG      PA      K      PBE
0.10859908 0.11566839 0.35260300 0.04290066 0.01990049 0.18886983 0.25325150
> L.r <- c(b.r, lam.r)
> L.r
      PKP      PP      IP      PLG      PA      K      PBE
0.10859908 0.11566839 0.35260300 0.04290066 0.01990049 0.18886983 0.25325150
0.03558539

```

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

Lampiran 9. Program algoritma Schall dan BIC

```
# function untuk train test split
train.test.split = function(df, prop.train=.8, seed=123) {
  set.seed(seed)
  smp_size <- floor(prop.train * nrow(df))
  train_ind <- sample(seq_len(nrow(df)), size = smp_size)
  split <- list("train"=df[train_ind, ], "test"=df[-train_ind, ])
  return(split)
}

# function untuk hitung MSE
calc.mse = function(y, y.hat){
  n = length(y)
  return(sum((y.hat - y)^2)/n)
}

# fungsi hitung MAPE
calc.mape = function(y, y.hat){
  n = length(y)
  return(sum(abs((y.hat - y)/y))/n)
}

#MENGHITUNG WAKTU EKSEKUSI

library(tictoc)

# fungsi untuk regresi ridge dengan algoritma Schall
ridge.with.Schall<- function(Y, X, lambda=Inf, converge=1e-4){
  n <- nrow(X)
  iot <- diag(ncol(X))
  dz <- Inf
  lambdas <- c()
  mse <-
  while(dz > converge){
    sol <- solve(t(X)%*%X+lambda*iot)
    beta <- sol%*%t(X)%*%Y
    hat <- X %*% sol%*%t(X)
    lam <- sum(diag(hat))/(t(beta)%*%beta)
    dz <- abs(lambda-lam)

    lambda<- as.numeric(lam)
  }
}
```

```

    lambdas = c(lambdas, lambda)
  }
  res <- Y-X%%beta
  logl <- .5* ( - n * (log(2 * pi) + 1 - log(n) +log(sum(res^2))))
  out <- list("coef"=round(beta,4), "logL"=logl, "hat"=hat, "lambdas"=lambdas,
"opt.lambda"=lambda)
  # out <- data.frame(beta, logl, hat, lambdas)
Out
}

ridge.with.ic <- function(Y,X,kappa)
{
  n <- nrow(X)
  iot <- diag(ncol(X))
  sol <- solve(t(X)%%X+kappa*iot)
  beta <- sol%%t(X)%%Y
  res <- Y-X%%beta
  #res <- apply(res, 2, crossprod)
  hat <- X %% sol%%t(X)
  logl <- .5*(-n*(log(2*pi)+log(sum((res)**2))-log(n) + 1))
  edf <- sum(diag(hat))
  Aic <- -2*logl+2*edf
  nm <- res/(1-edf/n)
  Bic <- -2*logl+log(n)*edf
  Gcv<- sum(nm^2)/n
  Aicc <- Aic + (2*edf^2+2*edf)/(n-edf-1)
  out <- list("coef"=round(beta,4),"logL"=logl, "hat"=hat,"AIC"=Aic,"Df"=edf, "RSS"=res,"GCV"=Gcv
    ,"BIC"=Bic,"lambda"=kappa,"AICc"=Aicc)
Out
}

# -----
# Persiapan dataset
# -----
dat<-read.csv("dataPDRB1.csv", sep = ",")
Dat
PDRB<-dat$PDRB
X<-dat[,-1]
ff<-function(x){(x-mean(x))/sd(x)}
Xx<-apply(X,2,ff)
Xx
df<-data.frame(PDRB,Xx)
Df

```

```

# train-test split
df.split <- train.test.split(df)
X.train <- as.matrix(df.split$train[,-1])
y.train <- df.split$train$PDRB
X.test <- as.matrix(df.split$test[,-1])
y.test <- df.split$test$PDRB

# =====

# -----
# Regresi ridge dengan algoritma Schall
# -----
tic("total")
# fit model ridge dengan algoritma Schall
fit.with.Schall<- ridge.with.Schall(y.train, X.train, lambda = 1,converge = 1e-10)
## lambda optimal
opt.lambda.Schall<- fit.with.Schall$opt.lambda
opt.lambda.Schall
## list lambda saat proses iterasi
lambdas.Schall<- fit.with.Schall$lambdas
lambdas.Schall
## plot lambda tiap iterasinya
plot(lambdas.Schall, col="orange", type="l", xlab="Iterasi", ylab="lambda")
points(lambdas.Schall)
## koefisien model
b.Schall<- fit.with.Schall$coef
b.Schall
## prediksi
y.predict.Schall<- X.test %*% b.Schall
y.predict.Schall
## MSE model
mse.Schall<- calc.mse(y.test, y.predict.Schall)
mse.Schall
## MAPE SCHALL
mape.Schall<- calc.mape(y.test, y.predict.Schall)
mape.Schall
toc()

# =====

# -----

```

```

# Regresi ridge dengan AIC dan BIC
# -----
tic("total")
# generate lambda
set.seed(123)
lambdas.generated <- 10^seq(-3, 5, length.out = 100)

# menghitung AIC – BIC
#aic <- c()
bic <- c()
for (lambda in lambdas.generated) {
  fit.with.ic <- ridge.with.ic(y.train, X.train, lambda)
  #aic <- c(aic, fit.with.ic$AIC)
  bic <- c(bic, fit.with.ic$BIC)
}

# plot IC terhadap lambda
plot(log(lambdas.generated), bic, col = "orange", type = "l", ylab = "Bayesian Information
Criterion")
#lines(log(lambdas_to_try), aic, col = "skyblue3")
#legend("bottomright", lwd = 1, col = c("orange", "skyblue3"), legend = c("AIC", "BIC"))

# lambda optimal
#lambda_aic <- lambdas.generated[which.min(aic)]
lambda_bic <- lambdas.generated[which.min(bic)]
lambda_bic
# fit final model dengan AIC dan BIC
#fit.with.aic <- ridge.with.ic(y.train, X.train, lambda_aic)
fit.with.bic <- ridge.with.ic(y.train, X.train, lambda_bic)
## koefisien model
#b.aic <- fit.with.aic$coef
b.bic <- fit.with.bic$coef
#b.aic
b.bic
## prediksi
#y.predict.aic <- X.test %*% b.aic
y.predict.bic <- X.test %*% b.bic
y.predict.bic
## MSE model
#mse.aic <- calc.mse(y.test, y.predict.aic)
mse.bic <- calc.mse(y.test, y.predict.bic)
#mse.aic
mse.bic

```

```
#  
#MAPE BIC  
mape.bic <- calc.mape(y.test, y.predict.bic)  
mape.bic  
#=====  
toc()
```



Lampiran 10. Output program algoritma Schall dan BIC

```

> # function untuk train test split
> train.test.split = function(df, prop.train=.8, seed=123) {

+   smp_size <- floor(prop.train * nrow(df))

+   split <- list("train"=df[train_ind, ], "test"=df[-train_ind, ])
+   return(split)
+ }
>
> # function untuk hitung MSE
> calc.mse = function(y, y.hat){

+   return(sum((y.hat - y)^2)/n)
+ }
> # fungsi hitung MAPE
> calc.mape = function(y, y.hat){

+   return(sum(abs((y.hat - y)/y))/n)
+ }
>
> #MENGHITUNG WAKTU EKSEKUSI
>
> library(tictoc)
>
>
> # fungsi untuk regresi ridge dengan algoritma Schall
> ridge.with.Schall<- function(Y, X, lambda=Inf, converge=1e-4){

+   lambdas <- c()
+   mse <-
+   while(dz > converge){
+     sol <- solve(t(X)%*%X+lambda*iot)
+     beta <- sol%*%t(X)%*%Y
+     hat <- X %*% sol%*%t(X)
+     lam <- sum(diag(hat))/(t(beta)%*%beta)
+     lambdas = c(lambdas, lambda)
+   }
+   res <- Y-X%*%beta
+   # out <- data.frame(beta, logl, hat, lambdas)
+ }

```

```

>
> ridge.with.ic <- function(Y,X,kappa)
+ {
+   sol <- solve(t(X)%*%X+kappa*iot)
+   beta <- sol%*%t(X)%*%Y
+   res <- Y-X%*%beta
+   #res <- apply(res, 2, crossprod)
+   hat <- X %*% sol%*%t(X)
+   logl <- .5*(-n*(log(2*pi)+log(sum((res)**2))-log(n) + 1))
+   out <- list("coef"=round(beta,4),"logL"=logl, "hat"=hat,"AIC"=Aic,"Df"=edf, "RSS"=res,"GCV"=Gcv
+     , "BIC"=Bic,"lambda"=kappa,"AICc"=Aicc)
+ }
>
>
> # =====
> # -----
> # Persiapan dataset
> # -----
> dat<-read.csv("dataPDRB1.csv", sep = ",")
> dat
  PDRB  PKP  PP  IP  PLG  PA  K  PBE
1  5.32 10.06 6.70 3.74 4.05 7.19 5.55 5.40
2  5.18 -3.49 7.77 5.04 5.52 6.10 7.08 8.07
3  5.17 -0.22 2.06 4.25 7.61 6.91 7.44 7.18
4  5.40  0.36 4.56 4.34 3.81 5.88 8.33 3.84
5  5.43  1.32 1.85 4.85 4.11 7.11 6.89 4.84
6  5.54  3.52 5.69 4.39 5.36 5.76 6.34 5.58
7  1.75 19.12 -0.25 -0.10 -3.12 1.39 -3.63 0.47
8  3.45  5.49 2.63 3.51 1.26 3.21 1.24 3.98
9  2.48  4.06 -2.10 1.88 2.55 -0.40 7.89 1.23
10 -2.35 -22.35 1.82 -0.09 5.42 0.69 0.16 -1.62
11 -0.84 11.76 2.63 -2.23 -4.44 -0.45 -7.60 -0.27
12 -5.18 11.76 -6.87 -3.95 -4.45 1.32 -5.70 -9.72
13  4.79  9.91 3.30 -0.99 7.93 0.62 8.16 6.58
14 -1.89 -21.65 -2.55 0.98 1.58 1.28 1.45 1.85
15  1.69 18.33 2.20 0.02 -0.04 0.99 -2.59 1.47
> PDRB<-dat$PDRB
> X<-dat[,-1]
> ff<-function(x){(x-mean(x))/sd(x)}
> Xx<-apply(X,2,ff)
> Xx
      PKP      PP      IP      PLG      PA      K

```

```
[1,] 0.56489064 1.25537227 0.71273875 0.39244561 1.36218170 0.5117848
[2,] -0.55067506 1.53891751 1.16902258 0.75911619 0.99252742 0.7898498
[3,] -0.28145736 0.02579290 0.89174241 1.28043695 1.26722464 0.8552768
[4,] -0.23370621 0.68828178 0.92333129 0.33258103 0.91791829 1.0170270
[5,] -0.15466982 -0.02985617 1.10233495 0.40741176 1.33505111 0.7553188
[6,] 0.02645524 0.98772676 0.94088067 0.71920647 0.87722241 0.6553608
[7,] 1.31079656 -0.58634683 -0.63505350 -1.39600886 -0.60478607 -1.1566047
[8,] 0.18864449 0.17684037 0.63201161 -0.30348019 0.01243485 -0.2715222
[9,] 0.07091320 -1.07658860 0.05990188 0.01829196 -1.21183302 0.9370606
[10,] -2.10341079 -0.03780603 -0.63154363 0.73417262 -0.84217873 -0.4678033
[11,] 0.70485091 0.17684037 -1.38265701 -1.72526408 -1.22878964 -1.8781195
[12,] 0.70485091 -2.34061739 -1.98635563 -1.72775843 -0.62852533 -1.5328101
[13,] 0.55254120 0.35438739 -0.94743243 1.36025640 -0.86591800 0.9861309
[14,] -2.04578009 -1.19583660 -0.25598693 -0.22366074 -0.64209063 -0.2333564
[15,] 1.24575620 0.06289228 -0.59293499 -0.62774669 -0.74043902 -0.9675932
```

PBE

```
[1,] 0.6301451
[2,] 1.2293215
[3,] 1.0295961
[4,] 0.2800645
[5,] 0.5044751
[6,] 0.6705390
[7,] -0.4761994
[8,] 0.3114820
[9,] -0.3056473
[10,] -0.9452177
[11,] -0.6422633
[12,] -2.7629439
[13,] 0.8949497
[14,] -0.1665127
[15,] -0.2517887
```

```
> df<-data.frame(PDRB,Xx)
```

```
> df
```

	PDRB	PKP	PP	IP	PLG	PA	K
1	5.32	0.56489064	1.25537227	0.71273875	0.39244561	1.36218170	0.5117848
2	5.18	-0.55067506	1.53891751	1.16902258	0.75911619	0.99252742	0.7898498
3	5.17	-0.28145736	0.02579290	0.89174241	1.28043695	1.26722464	0.8552768
4	5.40	-0.23370621	0.68828178	0.92333129	0.33258103	0.91791829	1.0170270
5	5.43	-0.15466982	-0.02985617	1.10233495	0.40741176	1.33505111	0.7553188
6	5.54	0.02645524	0.98772676	0.94088067	0.71920647	0.87722241	0.6553608
7	1.75	1.31079656	-0.58634683	-0.63505350	-1.39600886	-0.60478607	-1.1566047
8	3.45	0.18864449	0.17684037	0.63201161	-0.30348019	0.01243485	-0.2715222

```

9 2.48 0.07091320 -1.07658860 0.05990188 0.01829196 -1.21183302 0.9370606
10 -2.35 -2.10341079 -0.03780603 -0.63154363 0.73417262 -0.84217873 -0.4678033
11 -0.84 0.70485091 0.17684037 -1.38265701 -1.72526408 -1.22878964 -1.8781195
12 -5.18 0.70485091 -2.34061739 -1.98635563 -1.72775843 -0.62852533 -1.5328101
13 4.79 0.55254120 0.35438739 -0.94743243 1.36025640 -0.86591800 0.9861309
14 -1.89 -2.04578009 -1.19583660 -0.25598693 -0.22366074 -0.64209063 -0.2333564
15 1.69 1.24575620 0.06289228 -0.59293499 -0.62774669 -0.74043902 -0.9675932

```

PBE

```

1 0.6301451
2 1.2293215
3 1.0295961
4 0.2800645
5 0.5044751
6 0.6705390
7 -0.4761994
8 0.3114820
9 -0.3056473
10 -0.9452177
11 -0.6422633
12 -2.7629439
13 0.8949497
14 -0.1665127
15 -0.2517887

```

```

> # train-test split
> df.split <- train.test.split(df)
> X.train <- as.matrix(df.split$train[,-1])
> y.train <- df.split$train$PDRB
> X.test <- as.matrix(df.split$test[,-1])
> y.test <- df.split$test$PDRB
>
> # =====
>
> # -----
> # Regresi ridge dengan algoritma Schall
> # -----
> tic("total")
> # fit model ridge dengan algoritma Schall
> fit.with.Schall<- ridge.with.Schall(y.train, X.train, lambda = 1,converge = 1e-10)
> ## lambda optimal
> opt.lambda.Schall<- fit.with.Schall$opt.lambda
> opt.lambda.Schall
[1] 0.09405707

```

```

> ## list lambda saat proses iterasi
> lambdas.Schall<- fit.with.Schall$lambdas
> lambdas.Schall
[1] 0.37057546 0.20470914 0.14292945 0.11646202 0.10448286 0.09893956 0.09635010
[8] 0.09513538 0.09456445 0.09429588 0.09416948 0.09410999 0.09408198 0.09406880
[15] 0.09406259 0.09405967 0.09405829 0.09405764 0.09405734 0.09405719 0.09405713
[22] 0.09405709 0.09405708 0.09405707 0.09405707 0.09405707 0.09405707 0.09405707
[29] 0.09405707 0.09405707
> ## plot lambda tiap iterasinya
> plot(lambdas.Schall, col="orange", type="l", xlab="Iterasi", ylab="lambda")
> points(lambdas.Schall)
> ## koefisien model
> b.Schall<- fit.with.Schall$coef
> b.Schall
[,1]
PKP 0.0165
PP 0.6377
IP 5.7764
PLG -0.0479
PA 2.6839
K -5.1202
PBE -0.5460
> ## prediksi
> y.predict.Schall<- X.test %*% b.Schall
> y.predict.Schall
[,1]
8 5.034725
9 -8.223722
13 -13.164662
> ## MSE model
> mse.Schall<- calc.mse(y.test, y.predict.Schall)
> mse.Schall
[1] 146.4836
> ## MAPE SCHALL
> mape.Schall<- calc.mape(y.test, y.predict.Schall)
> mape.Schall
[1] 2.84124
> toc()
total: 0.64 sec elapsed
>
> # =====
>

```

```

>
> # -----
> # Regresi ridge dengan AIC dan BIC
> # -----
> tic("total")
> # generate lambda
> set.seed(123)
> lambdas.generated <- 10^seq(-3, 5, length.out = 100)
>
> # menghitung AIC – BIC
> #aic <- c()
> bic <- c()
> for (lambda in lambdas.generated) {

+ #aic <- c(aic, fit.with.ic$AIC)
+ bic <- c(bic, fit.with.ic$BIC)
+ }
>
> # plot IC terhadap lambda
> plot(log(lambdas.generated), bic, col = "orange", type = "l", ylab = "Bayesian Information
Criterion")
> #lines(log(lambdas_to_try), aic, col = "skyblue3")
> #legend("bottomright", lwd = 1, col = c("orange", "skyblue3"), legend = c("AIC", "BIC"))
>
> # lambda optimal
> #lambda_aic <- lambdas.generated[which.min(aic)]
> lambda_bic <- lambdas.generated[which.min(bic)]
> lambda_bic
[1] 0.02364489
> b.bic <- fit.with.bic$coef
> #b.aic
> b.bic
      [,1]
PKP 0.1157
PP 0.4526
IP 7.9048
PLG 0.3648
PA 2.3600
K -6.5579
PBE -1.2753
> ## prediksi
> #y.predict.aic <- X.test %*% b.aic
> y.predict.bic <- X.test %*% b.bic

```

```
> y.predict.bic
[1]
8 6.399809
9 -8.614158
13 -16.420561
> ## MSE model
> #mse.aic <- calc.mse(y.test, y.predict.aic)
> mse.bic <- calc.mse(y.test, y.predict.bic)
> #mse.aic
> mse.bic
[1] 193.8899
> #MAPE BIC
> mape.bic <- calc.mape(y.test, y.predict.bic)
> mape.bic
[1] 3.252187
> #=====
> toc()
total: 1.11 sec elapsed
```

Lampiran 11. Data PDRB Provinsi Jawa Tengah Menurut Lapangan Usaha

PDRB	PKP	PP	IP	PLG	PA	K	PBE
5.32	10.06	6.7	3.74	4.05	7.19	5.55	5.4
5.18	-3.49	7.77	5.04	5.52	6.1	7.08	8.07
5.17	-0.22	2.06	4.25	7.61	6.91	7.44	7.18
5.4	0.36	4.56	4.34	3.81	5.88	8.33	3.84
5.43	1.32	1.85	4.85	4.11	7.11	6.89	4.84
5.54	3.52	5.69	4.39	5.36	5.76	6.34	5.58
1.75	19.12	-0.25	-0.1	-3.12	1.39	-3.63	0.47
3.45	5.49	2.63	3.51	1.26	3.21	1.24	3.98
2.48	4.06	-2.1	1.88	2.55	-0.4	7.89	1.23
-2.35	-22.35	1.82	-0.09	5.42	0.69	0.16	-1.62
-0.84	11.76	2.63	-2.23	-4.44	-0.45	-7.6	-0.27
-5.18	11.76	-6.87	-3.95	-4.45	1.32	-5.7	-9.72
4.79	9.91	3.3	-0.99	7.93	0.62	8.16	6.58
-1.89	-21.65	-2.55	0.98	1.58	1.28	1.45	1.85
1.69	18.33	2.2	0.02	-0.04	0.99	-2.59	1.47

