

REGRESI KONTINUM MENGGUNAKAN PRAPEMROSESAN SELEKSI PEUBAH LASSO DAN SIR- LASSO PADA ALAT KELUARAN GLUKOSA DARAH NON- INVASIF

ADZKAR ADLU HASYR SURUDDIN



**PROGRAM STUDI STATISTIKA DAN SAINS DATA
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
BOGOR
2025**



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

PERNYATAAN MENGENAI TESIS DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa tesis dengan judul “Regresi Kontinum Menggunakan Prapemrosesan Seleksi Peubah LASSO dan SIR-LASSO pada Alat Keluaran Glukosa Darah Non-Invasif” adalah karya saya dengan arahan dari dosen pembimbing dan belum diajukan dalam bentuk apa pun kepada perguruan tinggi mana pun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka di bagian akhir tesis ini.

Dengan ini saya melimpahkan hak cipta dari karya tulis saya kepada Institut Pertanian Bogor.

Bogor, Juni 2025

Adzkar Adlu Hasyr Suruddin
G1501211033



RINGKASAN

ADZKAR ADLU HASYR SURUDDIN. Regresi Kontinum menggunakan Prapemrosesan Seleksi Peubah LASSO dan SIR-LASSO pada Alat Keluaran Glukosa Darah Non-Invasif. Dibimbing oleh ERFIANI dan I MADE SUMERTAJAYA.

Model kalibrasi merupakan metode yang mengaitkan hasil pengukuran dari alat atau sistem dengan standar yang diketahui atau nilai sebenarnya. Tujuan utama dari metode ini adalah menjamin keakuratan hasil pengukuran agar sesuai dengan nilai referensi yang diakui.

Multikolinearitas sering menjadi masalah dalam model kalibrasi, terutama saat menggunakan regresi linier berganda. Kondisi ini muncul ketika dua atau lebih peubah penjelas saling berkorelasi, sehingga sulit menilai pengaruh masing-masing peubah secara independen terhadap peubah respon.

Dampak utama dari multikolinearitas adalah ketidakstabilan koefisien regresi, di mana perubahan kecil pada data bisa menimbulkan perubahan besar pada koefisien. Situasi ini menyulitkan interpretasi model.

Regresi Kontinum (RK) dikembangkan dari metode RKT, RKU, dan RKTP sebagai solusi terhadap masalah multikolinearitas. Pendekatannya melibatkan proyeksi data asli ke peubah laten melalui matriks pembobot.

Prapemrosesan atau teknik pemadatan data menjadi langkah penting sebelum membangun model RK, khususnya ketika data memiliki dimensi besar atau saat jumlah peubah lebih banyak dibandingkan jumlah pengamatan. Penelitian ini menggunakan metode LASSO dan LASSO-SIR sebagai teknik prapemrosesan. LASSO bekerja dengan menyusutkan koefisien regresi untuk seleksi peubah, sedangkan SIR menggantikan peubah asli dengan kombinasi linier berdimensi rendah tanpa kehilangan informasi penting.

Studi ini membandingkan kinerja RK LASSO dan RK SIR-LASSO dalam memprediksi data berdimensi rendah maupun tinggi. Secara umum, RK LASSO menunjukkan performa prediksi yang lebih unggul, terutama melalui kesalahan prediksi yang lebih kecil.

Pada situasi dengan korelasi tinggi antar peubah penjelas, kedua metode mengalami penurunan performa akibat multikolinearitas. Kondisi ini mengakibatkan kesulitan dalam mengisolasi pengaruh masing-masing peubah secara akurat.

RK SIR-LASSO terbukti lebih unggul dibandingkan RK LASSO, khususnya saat jumlah peubah meningkat dan korelasi antar peubah. Nilai RMSEP yang lebih rendah menandakan efektivitas metode ini dalam menekan kesalahan prediksi.

Selain lebih akurat, RK SIR-LASSO menghasilkan prediksi yang lebih stabil, ditunjukkan dengan variabilitas kesalahan prediksi yang rendah. Keunggulan ini penting ketika berhadapan dengan data kompleks, khususnya yang melibatkan banyak peubah serta hubungan yang rumit antar peubah.

Penelitian ini mengungkapkan bahwa metode RK SIR-LASSO memiliki kinerja prediksi yang lebih unggul dibandingkan RK LASSO, terutama pada kondisi data berdimensi tinggi dengan korelasi antar peubah yang tinggi. Multikolinearitas menjadi faktor yang menyebabkan penurunan performa kedua



metode, namun RK SIR-LASSO tetap menunjukkan ketangguhannya dengan nilai RMSEP yang lebih rendah serta stabilitas prediksi yang lebih baik.

Penerapan pada data pengukuran kadar glukosa darah non-invasif tahun 2019 memperkuat temuan tersebut. RK SIR-LASSO lebih sesuai digunakan pada data tersebut, dengan jumlah slice optimal $H = 14$. Pemilihan jumlah slice yang tepat terbukti berpengaruh signifikan terhadap hasil prediksi. Jumlah slice yang terlalu kecil berpotensi menyebabkan hilangnya informasi penting, sedangkan jumlah yang terlalu besar dapat menurunkan kestabilan model. Hasil ini mendukung landasan teoritis sebelumnya mengenai pentingnya pemilihan jumlah slice yang optimal dalam meningkatkan akurasi dan konsistensi model prediksi.

Keywords: Glukosa Darah, LASSO, SIR-LASSO, Regresi Kontinum.



SUMMARY

ADZKAR ADLU HASYR SURUDDIN. The Continuum Regression with Preprocessed Selection Variable LASSO and SIR-LASSO on Non-Invasive Blood Glucose Output Device. Supervised by ERFIANI and I MADE SUMERTAJAYA.

A calibration model is a process that compares the measurement outputs of a system or instrument with a known standard or true value. The significant reason for this process is to make measurement results accurate so that they agree with reference values set.

Multicollinearity is a common issue in calibration models, particularly in multiple linear regression. It occurs when two or more of the predictor variables are related to one another, and hence it becomes hard to determine the independent contribution of each variable to the response variable.

The greatest effect of multicollinearity is the instability of regression coefficients in which the coefficients undergo drastic change with slight changes in the data. Such a situation makes interpretation of the model difficult.

Continuum Regression (CR) was formulated from the LSR, PCR, and PLSR methods as a method of solving the issue of multicollinearity. The method entails the transformation of the original data into latent variables through the use of a weighting matrix.

Preprocessing of the data or dimension reduction methods are significant procedures prior to constructing a CR model, particularly when the data dimension is large or the variable number exceeds the observation number. In this paper, LASSO and SIR-LASSO are employed as data preprocessing methods. LASSO conducts variable selection via shrinking regression coefficients, whereas SIR transforms the original variables into low-dimensional linear combinations without removing crucial information.

This paper contrasts the predictive capacity of CR LASSO and CR SIR-LASSO on low- and high-dimensional data. RK LASSO tends to perform better, with lower prediction errors.

Where predictor variables are highly correlated, both techniques are plagued by poor performance caused by multicollinearity. It is a troublesome condition because the effect of each variable cannot be precisely isolated.

CR SIR-LASSO was found to perform better than CR LASSO, particularly with large numbers of variables and high correlation among variables. The smaller RMSEP value reflects the effectiveness of this approach in minimizing prediction error.

Apart from being more precise, CR SIR-LASSO also generates more stable predictions, as indicated by low variability of prediction errors. An advantage such as this is particularly important when handling complex data, particularly those with a high number of variables and intricate variable relationships.

This research illustrates that CR SIR-LASSO has higher predictive ability than CR LASSO, particularly in the high-dimensional data context with high correlations between variables. Multicollinearity is among the reasons why the performance of both methods deteriorates, yet CR SIR-LASSO still shows its power with a lower RMSEP value and stabler prediction.

Application to 2019 non-invasive blood glucose measurement data confirms these results. CR SIR-LASSO is better adapted to deal with this data, for which the optimal number of slices is $H = 14$. The choice of sufficient number of slices was demonstrated to have great impacts on prediction performances. Too few slices can lead to losing valuable information, whereas too many slices can destabilize the model. These findings confirm the earlier theoretical foundation of the significance of choosing the best number of slices for enhancing the accuracy and reliability of prediction models.

Keywords: Blood Glucose, Continuum Regression, LASSO, SIR-LASSO

@Hak cipta milik IPB University

IPB University





Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

© Hak Cipta milik IPB, tahun 2025
Hak Cipta dilindungi Undang-Undang

Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan atau menyebutkan sumbernya. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik, atau tinjauan suatu masalah, dan pengutipan tersebut tidak merugikan kepentingan IPB.

Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apa pun tanpa izin IPB.

REGRESI KONTINUM MENGGUNAKAN PRAPEMROSESAN SELEKSI PEUBAH LASSO DAN SIR- LASSO PADA ALAT KELUARAN GLUKOSA DARAH NON- INVASIF

ADZKAR ADLU HASYR SURUDDIN

Tesis
sebagai salah satu syarat untuk memperoleh gelar
Magister Sains pada
Program Studi Statistika dan Sains Data

**PROGRAM STUDI STATISTIKA DAN SAINS DATA
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFROMATIKA
INSTITUT PERTANIAN BOGOR
BOGOR
2025**



Tim Penguji pada Ujian Tesis:
Dr. Septian Rahardiantoro, S.Stat., M.Si.

@Hak cipta milik IPB University

IPB University



IPB University
— Bogor Indonesia —

- Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

Judul Tesis : Regresi Kontinum menggunakan Prapemrosesan Seleksi Peubah
LASSO dan SIR-LASSO pada Alat Keluran Glukosa Darah
Non-Invasif

Nama : Adzkar Adlu Hasyr Suruddin

NIM : G1501211033

Disetujui oleh

Pembimbing 1:

Dr. Ir. Erfiani, M.Si.



Pembimbing 2:

Prof. Dr. Ir. I Made Sumertajaya, M.Si.



Diketahui oleh

Ketua Program Studi:

Dr. Agus Mohammad Soleh, S.Si., M.T.

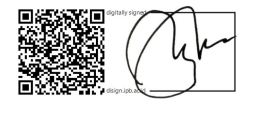
NIP 19750315 199903 1 004



Dekan Sekolah Sains Data, Matematika, dan Informatika:

Prof. Dr. Ir. Agus Buono, M.Si., M.Kom.

NIP 19660702 199302 1 001



Tanggal Ujian: 19 Februari 2025

Tanggal Lulus:



PRAKATA

Puji dan syukur penulis panjatkan kepada Allah subhanaahu wa ta'ala atas segala karunia-Nya sehingga karya ilmiah ini berhasil diselesaikan. Tema yang dipilih dalam penelitian yang dilaksanakan sejak bulan Maret 2023 sampai bulan Februari 2024 ini ialah model kalibrasi regresi kontinum menggunakan seleksi Peubah pada alat keluaran glukosa darah non-invasif, dengan judul “Regresi Kontinum menggunakan Prapemrosesan Seleksi Peubah LASSO dan SIR-LASSO pada Alat Keluaran Glukosa Darah Non-Invasif”.

Terima kasih penulis ucapkan kepada para pembimbing, Dr. Ir. Erfiani, M.Si. dan Prof. Dr. Ir. I Made Sumertajaya, M.Si. yang telah membimbing dan banyak memberi saran. Ucapan terima kasih juga disampaikan kepada moderator seminar dan penguji luar komisi pembimbing. Ungkapan terima kasih juga disampaikan kepada ayah, ibu dan kakak yang telah memberikan dukungan, doa, dan kasih sayangnya.

Semoga karya ilmiah ini bermanfaat bagi pihak yang membutuhkan dan bagi kemajuan ilmu pengetahuan.

Bogor, Juni 2025

Adzkar Adlu Hasyr Suruddin

DAFTAR ISI

DAFTAR TABEL	xii
DAFTAR GAMBAR	xii
DAFTAR LAMPIRAN	xii
II PENDAHULUAN	1
2.1 Latar Belakang	1
2.2 Tujuan	3
III TINJAUAN PUSTAKA	4
3.1 Regresi Kontinum	4
3.2 <i>Least Absolute Shrinkage and Selection Operator</i>	4
3.3 <i>Sliced Inverse Regression</i>	6
3.4 <i>Sliced Inverse Regression with LASSO</i>	7
3.5 Multikolinearitas	8
3.6 Ukuran Kebaikan Model	9
IV METODE	10
4.1 Data	10
4.2 Analisis Data	10
4.2.1 Data Simulasi	10
4.2.2 Data terapan	11
V HASIL DAN PEMBAHASAN	13
5.1 Kajian Data Simulasi	13
5.2 Pemilihan Jumlah <i>Slice</i> H	13
5.3 Hasil Simulasi RK menggunakan Prapemrosesan Seleksi Peubah LASSO dan SIR-LASSO	17
5.4 Penerapan Metode Pada Kadar Glukosa Darah Non-Invasif	21
5.4.1 Eksplorasi Data Kadar Glukosa Darah Invasif	21
5.4.2 Ekplorasi Data Kadar Glukosa Darah Non-Invasif.....	22
5.5 Kinerja Metode Regresi Kontinum	24
VI SIMPULAN DAN SARAN	26
6.1 Simpulan	26
6.2 Saran	26
DAFTAR PUSTAKA	27
LAMPIRAN.....	30
RIWAYAT HIDUP	34



DAFTAR TABEL

1	Tabel 1 Pemilihan jumlah H korelasi 0,5 p = 50	13
2	Tabel 2 Pemilihan jumlah H pada korelasi 0,5 dan p = 100	14
3	Tabel 3 Pemilihan jumlah H pada korelasi 0,5 dan p = 150	14
4	Tabel 4 Pemilihan jumlah H pada korelasi 0,9 dan p = 50	15
5	Tabel 5 Pemilihan jumlah H pada korelasi 0,9 dan p = 100	16
6	Tabel 6 Pemilihan jumlah H pada korelasi 0,9 dan p = 150	16
7	Tabel 7 Rata-rata ukuran kebaikan model RK	18
8	Tabel 8 Nilai VIF data kadar glukosa darah	24
9	Tabel 9 Evaluasi model RK	25

DAFTAR GAMBAR

10	Gambar 1 Ilustrasi metode LASSO	5
11	Gambar 2 Ilustrasi metode SIR	6
12	Gambar 3 Boxplot RMSEP perbandingan kedua metode	19
13	Gambar 4 Boxplot banyaknya peubah terseleksi	21
14	Gambar 5 Boxplot kadar glukosa darah invasif	22
15	Gambar 6 Output alat pengukuran kadar glukosa darah non-invasif	23

DAFTAR LAMPIRAN

16	Lampiran 1 Seleksi peubah menggunakan LASSO	31
17	Lampiran 2 Seleksi peubah menggunakan SIR-LASSO	32
18	Lampiran 3 Matriks Korelasi Kadar Glukosa Darah Invasif tahun 2019	33