

PENERAPAN *PRINCIPAL COMPONENT REGRESSION* DAN METODE PENYUSUTAN PADA PEMODELAN PREDIKSI TINGKAT PENGANGGURAN DI INDONESIA

IFTARR HENDRY



**PROGRAM STUDI MATEMATIKA
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
BOGOR
2026**



PERNYATAAN MENGENAI SKRIPSI DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa skripsi dengan judul “Penerapan *Principal Component Regression* dan Metode Penyusutan pada Pemodelan Prediksi Tingkat Pengangguran di Indonesia” 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 skripsi ini.

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

Bogor, Agustus 2026

Iftarr Hendry
G5401221073

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 mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



ABSTRAK

IFTARR HENDRY. Penerapan *Principal Component Regression* dan Metode Penyusutan pada Pemodelan Prediksi Tingkat Pengangguran di Indonesia. Dibimbing oleh Ngakan Komang Kutha Ardana dan Bidayatul Masulah.

Tingkat Pengangguran Terbuka (TPT) merupakan salah satu indikator penting yang mencerminkan kondisi pasar tenaga kerja dan perekonomian suatu negara. Pemodelan TPT sering melibatkan banyak variabel prediktor yang saling berkorelasi sehingga menimbulkan masalah multikolinearitas dan dapat menurunkan kestabilan model. Penelitian ini bertujuan membandingkan kinerja metode Regresi *Ridge*, *Least Absolute Shrinkage and Selection Operator* (LASSO), *Elastic Net*, dan *Principal Component Regression* (PCR) dalam memprediksi TPT di Indonesia. Data yang digunakan merupakan data tahunan Indonesia periode 1991–2024 yang terdiri atas satu variabel respon dan 21 variabel prediktor dari aspek ketenagakerjaan, ekonomi, demografi, pendidikan, perdagangan, dan investasi. Data dibagi menjadi data pelatihan dan pengujian dengan proporsi 70:30. Evaluasi model dilakukan menggunakan *Mean Absolute Error* (MAE), *Root Mean Squared Error* (RMSE), dan koefisien determinasi (R^2). Hasil penelitian menunjukkan bahwa model LASSO memberikan kinerja terbaik dengan MAE sebesar 0,2240, RMSE sebesar 0,3299, dan R^2 sebesar 0,9592. Pengujian Friedman menunjukkan adanya perbedaan kinerja yang signifikan antar model. Hasil ini menunjukkan bahwa LASSO merupakan metode yang paling efektif untuk memprediksi TPT pada data yang mengandung multikolinearitas.

Kata kunci: *Elastic Net*, LASSO, multikolinearitas, pengangguran terbuka, regresi *Ridge*

ABSTRACT

IFTARR HENDRY. Application of *Principal Component Regression* and *Shrinkage Methods* for Modeling Unemployment Rate Prediction in Indonesia. Supervised by Ngakan Komang Kutha Ardana and Bidayatul Masulah.

The Open Unemployment Rate (OUR) is a key indicator that reflects labor market and economic conditions within a country. Modeling the unemployment rate often involves numerous predictor variables that are highly correlated, leading to multicollinearity problems and reducing model stability. This study aims to compare the performance of *Ridge Regression*, *Least Absolute Shrinkage and Selection Operator* (LASSO), *Elastic Net*, and *Principal Component Regression* (PCR) in predicting the unemployment rate in Indonesia. The study utilized annual Indonesian data from 1991 to 2024, consisting of one response variable and 21 predictor variables that represented labor, economic, demographic, educational, trade, and investment aspects. The data were divided into training and Testing sets using a 70:30 proportion. Model performance was evaluated using *Mean Absolute Error* (MAE), *Root Mean Squared Error* (RMSE), and the coefficient of determination (R^2). The results showed that the LASSO model achieved the best predictive performance with a MAE of 0.2240, an RMSE of 0.3299, and a R^2 of 0.9592. Friedman's *Test* indicated a significant difference in performance among

the models. These findings suggest that LASSO is the most effective method for predicting the unemployment rate in the presence of multicollinearity.

Keywords: *Elastic Net, LASSO, multicollinearity, open unemployment rate, Ridge regression*

@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 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 2026
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.

PENERAPAN *PRINCIPAL COMPONENT REGRESSION* DAN METODE PENYUSUTAN PADA PEMODELAN PREDIKSI TINGKAT PENGANGGURAN DI INDONESIA

IFTARR HENDRY

Skripsi
sebagai salah satu syarat untuk memperoleh gelar
Sarjana Matematika pada
Program Studi Matematika

**PROGRAM STUDI MATEMATIKA
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
BOGOR
2026**



@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.

Tim Penguji pada Ujian Skripsi:

1. Prof. Dr. Ir. I Wayan Mangku, M.Sc.

Perpustakaan IPB University



Judul Skripsi : Penerapan *Principal Component Regression* dan Metode
Penyusutan pada Pemodelan Prediksi Tingkat Pengangguran di
Indonesia

Nama : Iftarr Hendry
NIM : G5401221073

@Hak cipta milik IPB University

Disetujui oleh

Pembimbing 1:

Ir. Ngakan Komang Kutha Ardana, M.Sc

Pembimbing 2:

Bidayatul Masulah S.Si., M.Sc

Diketahui oleh

Ketua Program Studi:

Dr. Donny Citra Lesmana, S.Si., M.Fin.Math.
197902272005011001

Tanggal Ujian:
21 Juli 2026

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 Oktober 2025 sampai bulan Juli 2026 ini ialah tugas akhir dengan judul “Penerapan *Principal Component Regression* dan Metode Penyusutan pada Pemodelan Prediksi Tingkat Pengangguran di Indonesia”. Dalam penyusunan karya ilmiah ini, penulis menyadari keberhasilan tidak terlepas dari bantuan, dukungan dan doa dari berbagai pihak. Oleh karena itu pada kesempatan ini penulis ingin menyampaikan terima kasih sebesar-besarnya kepada

1. Kedua orang tua penulis yaitu Bapak Hendry dan Ibu Zinnia Afriediza beserta para saudara penulis atas segala doa, kasih sayang, dukungan, motivasi, serta pengorbanan yang tiada henti diberikan kepada penulis.
2. Bapak Ir. Ngakan Komang Kutha Ardana, M.Sc. selaku Dosen Pembimbing I dan Ibu Bidayatul Masulah S.Si., M.Sc selaku Dosen Pembimbing II yang telah dengan sabar memberikan bimbingan, arahan, motivasi, serta masukan selama proses penyusunan karya ilmiah ini.
3. Bapak Prof. Dr. Ir. I Wayan Mangku, M.Sc selaku Dosen Penguji yang telah memberikan dedikasinya dalam memberi arahan serta masukan kepada penulis dalam menyelesaikan tugas akhir.
4. Seluruh dosen dan tenaga kependidikan Program Studi Matematika SSMI IPB University yang telah memberikan ilmu dan nasihat selama penulis menjadi mahasiswa di IPB University.
5. Arya, Galuh dan Mualim selaku Founder Arif Jaya yang selalu mendukung dan memotivasi penulis dalam menyelesaikan studi di Departemen Matematika SSMI IPB.
6. Salina, pipip, para kucing dan ayam peliharaan lain yang selalu membuat penulis merasa semangat dan bahagia selama menyusun karya tulis.
7. Seluruh mahasiswa angkatan 59 Program Studi Matematika SSMI IPB University dan sahabat terdekat dalam PKU ST 24 yang telah menjadi teman seperjuangan penulis selama di IPB University

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

Bogor, Agustus 2026

Iftarr Hendry



DAFTAR ISI

DAFTAR TABEL	xi
DAFTAR GAMBAR	xi
DAFTAR LAMPIRAN	xii
PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Tujuan	3
TINJAUAN PUSTAKA	4
2.1 Pengangguran	4
2.2 Regresi Linear	4
2.3 <i>Principal Component Analysis</i>	5
2.4 <i>Principal Components Regression</i>	7
2.5 Regresi Ridge	8
2.6 Regresi LASSO	8
2.7 Regresi <i>Elastic Net</i>	9
2.8 Validasi Silang	10
2.9 Ukuran Evaluasi Model	10
2.10 Analisis Perbandingan Kinerja Model	12
III METODE	14
3.1 Data	14
3.2 Tahapan Penelitian	15
IV HASIL DAN PEMBAHASAN	18
4.1 Statistik Deskriptif	18
4.2 Analisis Korelasi dan Multikolinearitas	19
4.3 <i>Principal Component Analysis</i>	22
4.4 Pemodelan Metode Penyusutan	24
4.5 <i>Principal Component Regression</i>	33
4.6 Perbandingan Evaluasi Model	35
4.7 Uji Perbedaan Kinerja Model	36
V SIMPULAN DAN SARAN	41
5.1 Simpulan	41
5.2 Saran	42
DAFTAR PUSTAKA	43
LAMPIRAN	45
RIWAYAT HIDUP	52

DAFTAR TABEL

1	Deskripsi lengkap variabel penelitian	14
2	Nilai <i>variance inflation factor</i> variabel prediktor	20
3	Nilai eigen, proporsi keragaman, dan keragaman kumulatif komponen utama	22
4	Hasil evaluasi model regresi <i>ridge</i>	25
5	Koefisien model regresi <i>ridge</i> pada λ optimum	26
6	Hasil evaluasi model regresi LASSO	28
7	Koefisien model regresi <i>elastic net</i> pada λ optimum	31
8	Hasil evaluasi model regresi <i>elastic net</i>	31
9	Perbandingan kinerja model regresi <i>ridge</i> , LASSO, dan <i>elastic net</i> pada data pengujian	32
10	Estimasi parameter model <i>principal component regression</i>	33
11	Hasil evaluasi model <i>principal component regression</i> pada data uji	34
12	Hasil evaluasi empat model prediksi pada data pengujian	35
13	Hipotesis uji Friedman untuk perbandingan kinerja model	37
14	Hasil uji Friedman untuk menguji perbedaan kinerja model	37
15	Hasil uji lanjut (<i>post hoc</i>) Nemenyi antar model prediksi	38
16	Hasil uji normalitas shapiro-wilk pada <i>absolute error</i> model	38
17	Hasil Wilcoxon <i>signed-rank test</i> pada <i>absolute error</i> antar model prediksi	40

DAFTAR GAMBAR

1	Diagram alir tahapan penelitian	17
2	Boxplot distribusi variabel penelitian berdasarkan kelompok skala data: (a) kelompok I, (b) kelompok II, (c) kelompok III, dan (d) variabel berskala besar setelah transformasi logaritma basis 10.	18
3	Visualisasi korelasi antar variabel prediktor menggunakan heatmap	20
4	Hasil analisis komponen utama (a) Scree plot proporsi variansi yang dijelaskan setiap komponen utama (b) plot variansi kumulatif komponen utama	23
5	Kurva <i>cross validation</i> untuk penentuan parameter penalti (λ) pada regresi <i>ridge</i>	24
6	Kurva <i>cross validation</i> untuk penentuan parameter penalti (λ) pada regresi LASSO	27
7	Jalur koefisien regresi LASSO terhadap perubahan nilai λ	27
8	Jalur koefisien regresi <i>Elastic Net</i> terhadap perubahan nilai λ	30
9	Kurva <i>cross validation</i> untuk penentuan parameter penalti (λ) pada regresi <i>elastic net</i>	30
10	Perbandingan kinerja model pada data pengujian berdasarkan <i>Root Mean Squared Error</i>	35

11	Normal Q-Q Plot <i>Absolute Error</i> empat model prediksi: (a) Regresi Ridge, (b) LASSO, (c) <i>Elastic Net</i> , dan (d) <i>Principal Component Regression</i>	39
12	Perbandingan sebaran <i>absolute error</i> empat model prediksi pada data pengujian	39

DAFTAR LAMPIRAN

1	Variasi data pengujian	46
2	<i>Syntax</i> import data dan definisi X dan Y	46
3	<i>Syntax</i> split data latih dan data uji	47
4	<i>Syntax</i> cek VIF dan multikolinearitas	47
5	<i>Syntax</i> regresi <i>ridge</i>	48
6	<i>Syntax</i> regresi LASSO	48
7	<i>Syntax</i> regresi <i>elastic net</i>	49
8	<i>Syntax</i> pembentukan PCA	49
9	<i>Syntax</i> PCR	50
10	<i>Syntax</i> uji Friedman	50
11	<i>Syntax post hoc</i> Nemenyi	50
12	<i>Syntax</i> analisis perbedaan <i>absolute error</i> antar model	51