

PERBANDINGAN METODE *ELASTIC NET PRINCIPAL COMPONENT REGRESSION* (ENPCR) DAN *SPARSE PRINCIPAL COMPONENT REGRESSION* (SPCR) DALAM MENGATASI MULTIKOLINEARITAS

(Skripsi)

Oleh

DEWI INTAN NUR'AINI



**JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2025**

ABSTRACT

COMPARISON OF ELASTIC NET PRINCIPAL COMPONENT REGRESSION (ENPCR) AND SPARSE PRINCIPAL COMPONENT REGRESSION (SPCR) METHODS IN ADDRESSING MULTICOLLINEARITY

By

Dewi Intan Nur'aini

Elastic Net Principal Component Regression and Sparse Principal Component Regression are methods used to address multicollinearity in regression models by incorporating dimensionality reduction and regularization techniques. Multicollinearity occurs when independent variables have high correlations, which can lead to unstable and unreliable parameter estimates. The objective of this study is to estimate regression parameters using the Elastic Net Principal Component Regression and Sparse Principal Component Regression methods to determine the best method for handling multicollinearity. The results of this study show that the Elastic Net Principal Component Regression method is more effective in handling multicollinearity compared to the Sparse Principal Component Regression method, based on the values of MMAD and MSE.

Keywords: multicollinearity, elastic net principal component regression, sparse principal component regression, MMAD, MSE.

ABSTRAK

PERBANDINGAN METODE *ELASTIC NET PRINCIPAL COMPONENT REGRESSION* (ENPCR) DAN *SPARSE PRINCIPAL COMPONENT REGRESSION* (SPCR) DALAM MENGATASI MULTIKOLINEARITAS

Oleh

Dewi Intan Nur'aini

Elastic Net Principal Component Regression dan *Sparse Principal Component Regression* adalah metode yang digunakan untuk mengatasi multikolinearitas dalam model regresi dengan menggabungkan teknik reduksi dimensi dan regularisasi. Multikolinearitas terjadi ketika variabel independen memiliki korelasi tinggi, yang dapat menyebabkan estimasi parameter menjadi tidak stabil dan kurang dapat diandalkan. Tujuan dari penelitian ini adalah menduga parameter regresi dengan menggunakan metode *Elastic Net Principal Component Regression* dan *Sparse Principal Component Regression* untuk mengetahui metode terbaik dalam menangani multikolinearitas. Hasil dari penelitian ini menunjukkan bahwa metode *Elastic Net Principal Component Regression* lebih baik dalam menangani multikolinearitas dibandingkan metode *Sparse Principal Component Regression* berdasarkan nilai dari MMAD dan MSE.

Kata kunci : multikolinearitas, *elastic net principal component regression*, *sparse principal component regression*, MMAD, MSE.

PERBANDINGAN METODE *ELASTIC NET PRINCIPAL COMPONENT REGRESSION* (ENPCR) DAN *SPARSE PRINCIPAL COMPONENT REGRESSION* (SPCR) DALAM MENGATASI MULTIKOLINEARITAS

Oleh

DEWI INTAN NUR'AINI

SKRIPSI

**Sebagai Salah Satu Syarat untuk Memperoleh Gelar
SARJANA MATEMATIKA**

Pada

**Jurusan Matematika
Fakultas Matematika dan Ilmu Pengetahuan
Universitas lampung**



**JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2025**

Judul : **PERBANDINGAN METODE *ELASTIC NET*
PRINCIPAL COMPONENT REGRESSION (ENPCR)
DAN *SPARSE PRINCIPAL COMPONENT*
REGRESSION (SPCR) DALAM MENGATASI
MULTIKOLINEARITAS**

Nama Mahasiswa : **Dewi Intan Nur'aini**

NPM : **2117031012**

Program Studi : **Matematika**

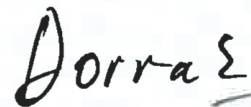
Fakultas : **Matematika dan Ilmu Pengetahuan Alam**

MENYETUJUI,

1. Komisi Pembimbing

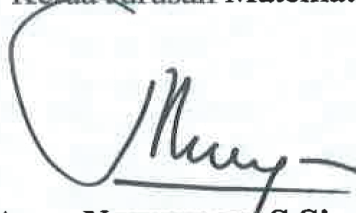


Prof. Ir. Netti Herawati, M.Sc., Ph.D.
NIP. 196501251990032001



Dra. Dorrah Azis, M.Si.
NIP. 196101281988112001

2. Ketua Jurusan Matematika



Dr. Aang Nuryaman, S.Si., M.Si.
NIP. 197403162005011001

MENGESAHKAN

1. Tim Penguji

Ketua

: **Prof. Ir. Netti Herawati, M.Sc., Ph.D.**



Sekretaris

: **Dra. Dorrah Azis, M.Si.**



Penguji

Bukan Pembimbing

: **Dr. Khoirin Nisa, S.Si., M.Si.**



2. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam



Dr. Eng. Heri Satria, S.Si., M.Si.

NIP. 197110012005011002

Tanggal Lulus Ujian Skripsi: **10 Maret 2025**

PERNYATAAN SKRIPSI MAHASISWA

Yang bertanda tangan di bawah ini:

Nama Mahasiswa : **Dewi Intan Nur'aini**
Nomor Pokok Mahasiswa : **2117031012**
Jurusan : **Matematika**
Judul Skripsi : **PERBANDINGAN METODE *ELASTIC NET*
PRINCIPAL COMPONENT REGRESSION
(ENPCR) DAN *SPARSE PRINCIPAL*
COMPONENT REGRESSION (SPCR) DALAM
MENGATASI MULTIKOLINEARITAS**

Dengan ini menyatakan bahwa penelitian ini adalah hasil kerja saya sendiri.
Apabila dikemudian hari terbukti bahwa skripsi ini merupakan salinan atau dibuat
oleh orang lain, maka saya bersedia menerima sanksi sesuai dengan ketentuan
akademik yang berlaku.

Bandar Lampung, 10 Maret 2025

Penulis,



Dewi Intan Nur'aini
NPM. 2117031012

RIWAYAT HIDUP

Penulis Dewi Intan Nur'aini dilahirkan di Gajah Timur III pada 8 Oktober 2002. Penulis merupakan anak kedua dari empat bersaudara dari pasangan Bapak Awi dan Ibu Nur Syamsiah.

Penulis pertama kali menempuh pendidikannya di Raudhatul Athfal Khodijah Rejo Asri 3 pada tahun 2008-2009, kemudian melanjutkan pendidikannya ke Sekolah Dasar di SD Negeri 02 Rejo Asri 3 dari tahun 2009 hingga 2015. Setelah menyelesaikan pendidikan dasar, penulis melanjutkan pendidikannya di SMP Negeri 02 Kotagajah pada tahun 2015-2018, lalu melanjutkan ke SMA Negeri 01 Kotagajah pada tahun 2018-2021. Kemudian pada tahun 2021 penulis diterima sebagai mahasiswa Program Studi S1 Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lampung melalui jalur Seleksi Nasional Masuk Perguruan Tinggi Negeri (SNMPTN).

Selama menjalani pendidikan di Universitas Lampung, penulis aktif dalam organisasi. Pada tahun 2023, penulis menjadi bagian dari Badan Eksekutif Mahasiswa (BEM) FMIPA UNILA. Pada tahun 2024, penulis melaksanakan Kerja Praktik (KP) di Dinas Perhubungan Kota Bandar Lampung dan sebagai salah satu bentuk pengabdian kepada masyarakat penulis melaksanakan Kuliah Kerja Nyata (KKN) di Desa Gedung Wani Timur, Kecamatan Marga Tiga, Kabupaten Lampung Timur.

KATA INSPIRASI

*"Janganlah kamu bersedih, sesungguhnya Allah bersama kita."
(QS. At-Taubah: 40)*

*"Allah tidak membebani seseorang melainkan sesuai dengan kesanggupannya."
(QS. Al-Baqarah: 286)*

*"Barang siapa yang menempuh jalan untuk menuntut ilmu, niscaya Allah subhanahu
wata'ala akan memudahkan baginya jalan menuju surga."
(H.R. Muslim)*

*"Orang yang memindahkan gunung memulai dengan membawa batu-batu kecil."
(Konfusius)*

*"Hidup itu seperti mengendarai sepeda. Untuk menjaga keseimbangan, kamu harus
terus bergerak."
(Albert Einstein)*

*"Semua jatuh bangunmu hal yang biasa, angan dan pertanyaan waktu yang
menjawabnya, berikan tenggat waktu bersedihlah secukupnya, rayakan perasaanmu
sebagai manusia"
(Hindia)*

*"Mimpi yang kau genggam bukan sekadar tujuan, tetapi bukti bahwa kau berani
melawan ketidakpastian."
(Dewi Intan Nur'aini)*

PERSEMBAHAN

Puji dan syukur kepada Allah SWT yang telah memberikan petunjuk, kemudahan, serta kelancaran dalam menyelesaikan skripsi ini. Dengan penuh rasa syukur dan kebahagiaan saya persembahkan karya sederhana ini kepada:

Orang Tua, Kakak dan Adik Tercinta

Yang senantiasa memberikan cinta, doa, dan dukungan tanpa henti, menjadi sumber kekuatan dan semangat bagi penulis dalam menyelesaikan skripsi ini. Tanpa kalian, perjalanan ini tidak akan sampai pada titik ini. Setiap langkah penuh dengan perjuangan dan cerita, yang tak lepas dari doa tulus Mamah, Bapak, Mbak Puji, Mas Bari, Rifqi, dan Akbar. Terima kasih atas kasih sayang, pengorbanan, serta kepercayaan yang tiada henti kalian berikan.

Dosen Pembimbing dan Penguji

Yang tulus memberikan motivasi, ilmu, dan arahan yang berharga bagi penulis

Almamater Tercinta, Universitas Lampung

SANWACANA

Alhamdulillah Rabbil ‘Alamin, segala puji dan syukur penulis panjatkan ke hadirat Allah SWT atas limpahan rahmat, taufik, dan hidayah-Nya, sehingga penulis dapat menyelesaikan skripsi dengan judul “Perbandingan Metode *Elastic Net Principal Component Regression* (ENPCR) dan *Sparse Principal Component Regression* (SPCR) Dalam Mengatasi Multikolinearitas”.

Penyusunan skripsi ini tidak akan terselesaikan dengan baik tanpa adanya bantuan, bimbingan dan dukungan dari berbagai pihak. Sehingga dalam kesempatan ini penulis ingin mengucapkan terima kasih kepada:

1. Ibu Prof. Ir. Netti Herawati, M.Sc., Ph.D., selaku dosen pembimbing I dan dosen pembimbing akademik yang telah memberikan bimbingan, arahan dan masukan dalam proses menyelesaikan skripsi.
2. Ibu Dra. Dorrah Azis, M.Si., selaku dosen pembimbing II yang telah memberikan bimbingan, arahan, serta saran dalam proses penyusunan skripsi.
3. Ibu Dr. Khoirin Nisa, S.Si., M.Si., selaku dosen penguji yang telah memberikan kritik dan saran, sehingga skripsi ini dapat terselesaikan.
4. Bapak Dr. Aang Nuryaman, S.Si, M.Si., selaku Ketua Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
5. Seluruh dosen Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
6. Bapak Dr. Eng. Heri Satria, S.Si., M.Si. selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
7. Orang tua tercinta, Bapak Awi dan Ibu Nur Syamsiah. Terima kasih atas segala pengorbanan dan perjuangan dalam memberikan kehidupan yang layak bagi penulis, serta atas kasih sayang tulus yang diberikan. Dengan segala

keterbatasan, mereka tetap memberikan yang terbaik dan tak kenal lelah mendoakan serta memberikan perhatian dan dukungan hingga penulis mampu menyelesaikan studinya sampai meraih gelar sarjana.

8. Ketiga saudara kandung penulis, Mbak Puji, Rifqi, dan Akbar. Terima kasih atas kebersamaan, dukungan, dan kasih sayang yang tak ternilai. Kalian selalu ada di setiap langkah perjalanan ini, dari hal-hal sederhana hingga momen-momen yang paling berharga.
9. Kepada Kakak ipar Sobari dan keponakan Haura, terima kasih atas kasih sayang dan dukungan yang telah kalian berikan kepada penulis. Kehadiran kalian membawa kebahagiaan dan kehangatan, menjadikan setiap langkah dalam perjalanan ini lebih bermakna.
10. Kepada Keluarga Besar Kakek Abu Rosid, dan Keluarga Besar Alm. Kakek Dul Hadi. Terima Kasih atas doa dan dukungan yang senantiasa diberikan.
11. Rhea, Lola, Salma, dan Shelvi. Terima kasih atas kebersamaan yang penuh warna selama masa perkuliahan. Setiap tawa, cerita, dan dukungan yang kalian berikan telah menjadi bagian berharga dalam perjalanan ini. *Good luck!*
12. Teman – teman Jurusan Matematika FMIPA UNILA angkatan 2021.
13. Bella, Laras, Putri, Arisa dan semua teman satu Pondokan Sakura. Terima kasih atas setiap momen yang telah kita jalani bersama. Kehangatan, kebersamaan, dan tawa yang kalian hadirkan menjadikan masa-masa ini lebih berkesan. Semoga persahabatan ini tetap terjaga dan membawa kita menuju masa depan yang gemilang.
14. Teman sejak kecil, Livia, Latifa, dan Masitoh. Kalian adalah bagian tak tergantikan dalam perjalanan hidup penulis. Terima kasih atas setiap tawa, cerita, dan dukungan yang telah kalian berikan. Persahabatan kita adalah salah satu anugerah terindah yang selalu penulis syukuri.
15. Sahabat penulis, Elisa. Terima kasih telah menjadi tempat berbagi di setiap suka dan duka. Kesabaranmu dalam mendengarkan keluh kesah, nasihat yang penuh makna, serta dukungan dan kepercayaan yang selalu kamu berikan menjadi sumber kekuatan bagi penulis. Persahabatanmu adalah hadiah berharga yang tak tergantikan, dan penulis selalu bersyukur akan itu.

16. Seluruh pihak terkait yang tidak bisa penulis sebutkan satu persatu yang telah terlibat dalam menyelesaikan skripsi ini.
17. Terakhir, Dewi Intan Nur'aini, ya! Diri saya sendiri. Apresiasi sebesar-besarnya yang telah berjuang menyelesaikan apa yang telah dimulai. Sulit bisa bertahan sampai titik ini, terima kasih untuk tetap hidup dan merayakan dirimu sendiri, walaupun sering kali putus asa atas apa yang sedang diusahakan. Tetaplah jadi manusia yang mau berusaha dan tidak lelah untuk mencoba. *God, thank You for helping me become an independent woman. I know there are many great ones out there, but I am proud of what I have achieved.*

Penulis menyadari bahwa dalam penulisan skripsi ini masih jauh dari sempurna, sehingga informasi tambahan, saran, dan kritik untuk pengembangan lebih lanjut sangat diharapkan. Semoga skripsi ini dapat bermanfaat bagi kita semua.

Bandar Lampung, 10 Maret 2025
Penulis,

Dewi Intan Nur'aini

DAFTAR ISI

Halaman

DAFTAR TABEL	iii
---------------------------	------------

DAFTAR GAMBAR.....	vi
---------------------------	-----------

I. PENDAHULUAN	1
-----------------------------	----------

1.1 Latar Belakang dan Masalah	1
--------------------------------------	---

1.2 Tujuan Penelitian.....	3
----------------------------	---

1.3 Manfaat Penelitian.....	4
-----------------------------	---

II. TINJAUAN PUSTAKA.....	5
----------------------------------	----------

2.1 Analisis Regresi	5
----------------------------	---

2.2 Analisis Regresi Linear Berganda	5
--	---

2.3 Asumsi Analisis Regresi Linear	7
--	---

2.4 Metode Kuadrat Terkecil (MKT)	8
---	---

2.5 Multikolinearitas.....	10
----------------------------	----

2.6 Standarisasi Variabel	11
---------------------------------	----

2.7 <i>Principal Component Analysis</i> (PCA).....	14
--	----

2.8 <i>Principal Component Regression</i> (PCR).....	16
--	----

2.9 <i>Elastic Net Principal Component Regression</i> (ENPCR).....	19
--	----

2.10 <i>Sparse Principal Component Regression</i> (SPCR).....	20
---	----

2.11 Ketepatan Model.....	24
---------------------------	----

III. METODOLOGI PENELITIAN	26
---	-----------

3.1 Tempat dan Waktu Penelitian.....	26
--------------------------------------	----

3.2 Data Penelitian.....	26
--------------------------	----

3.3	Metode Penelitian	27
IV.	HASIL DAN PEMBAHASAN	29
4.1	Hasil Simulasi untuk Kelompok Data dengan $n = 100$ dan $\rho = 0.90$..	29
4.2	Hasil Simulasi untuk Kelompok Data dengan $n = 100$ dan $\rho = 0.99$...	35
4.3	Hasil Simulasi untuk Kelompok Data dengan $n = 200$ dan $\rho = 0.90$...	41
4.4	Hasil Simulasi untuk Kelompok Data dengan $n = 200$ dan $\rho = 0.99$...	47
4.5	Hasil Simulasi untuk Kelompok Data dengan $n = 300$ dan $\rho = 0.90$...	53
4.6	Hasil Simulasi untuk Kelompok Data dengan $n = 300$ dan $\rho = 0.99$...	59
4.7	Perbandingan MMAD dan MSE Metode ENPCR dan SPCR.....	65
V.	KESIMPULAN	69
	DAFTAR PUSTAKA	70
	LAMPIRAN	

DAFTAR TABEL

Tabel	Halaman
1. Tingkat Korelasi antar variabel Prediktor dengan $n = 100$ dan $\rho = 0.90$	29
2. Nilai VIF Variabel Prediktor dengan $n = 100$ dan $\rho = 0.90$	30
3. Nilai eigen dan Vektor eigen dengan $n = 100$ dan $\rho = 0.90$	31
4. Presentase Keragaman Komponen Utama dan Nilai eigen dengan $n = 100$ dan $\rho = 0.90$	32
5. Loadings SPCR Data $n = 100$ dan $\rho = 0.90$	33
6. Nilai Rata-rata Penduga Parameter, MMAD dan MSE dengan $n = 100$ dan $\rho = 0.90$	34
7. Tingkat Korelasi antar variabel Prediktor dengan $n = 100$ dan $\rho = 0.99$	35
8. Nilai VIF Variabel Prediktor dengan $n = 100$ dan $\rho = 0.99$	36
9. Nilai eigen dan Vektor eigen dengan $n = 100$ dan $\rho = 0.99$	37
10. Presentase Keragaman Komponen Utama dan Nilai eigen dengan $n = 100$ dan $\rho = 0.99$	38
11. Loadings SPCR Data $n = 100$ dan $\rho = 0.99$	39
12. Nilai Rata-rata Penduga Parameter, MMAD dan MSE dengan $n = 100$ dan $\rho = 0.99$	40
13. Tingkat Korelasi antar variabel Prediktor dengan $n = 200$ dan $\rho = 0.90$	41

14. Nilai VIF Variabel Prediktor dengan $n = 200$ dan $\rho = 0.90$	42
15. Nilai eigen dan Vektor eigen dengan $n = 200$ dan $\rho = 0.90$	43
16. Presentase Keragaman Komponen Utama dan Nilai eigen dengan $n = 200$ dan $\rho = 0.90$	44
17. Loadings SPCR Data $n = 200$ dan $\rho = 0.90$	45
18. Nilai Rata-rata Penduga Parameter, MMAD dan MSE dengan $n = 200$ dan $\rho = 0.90$	46
19. Tingkat Korelasi antar variabel Prediktor dengan $n = 200$ dan $\rho = 0.99$	47
20. Nilai VIF Variabel Prediktor dengan $n = 200$ dan $\rho = 0.99$	48
21. Nilai eigen dan Vektor eigen dengan $n = 200$ dan $\rho = 0.99$	49
22. Presentase Keragaman Komponen Utama dan Nilai eigen dengan $n = 200$ dan $\rho = 0.99$	50
23. Loadings SPCR Data $n = 200$ dan $\rho = 0.99$	51
24. Nilai Rata-rata Penduga Parameter, MMAD dan MSE dengan $n = 200$ dan $\rho = 0.99$	52
25. Tingkat Korelasi antar variabel Prediktor dengan $n = 300$ dan $\rho = 0.90$	53
26. Nilai VIF Variabel Prediktor dengan $n = 300$ dan $\rho = 0.90$	54
27. Nilai eigen dan Vektor eigen dengan $n = 300$ dan $\rho = 0.90$	55
28. Presentase Keragaman Komponen Utama dan Nilai eigen dengan $n = 300$ dan $\rho = 0.90$	56
29. Loadings SPCR Data $n = 300$ dan $\rho = 0.90$	57
30. Nilai Rata-rata Penduga Parameter, MMAD dan MSE dengan $n = 300$ dan $\rho = 0.90$	58

31. Tingkat Korelasi antar variabel Prediktor dengan $n = 300$ dan $\rho = 0.99$	59
32. Nilai VIF Variabel Prediktor dengan $n = 300$ dan $\rho = 0.99$	60
33. Nilai eigen dan Vektor eigen dengan $n = 300$ dan $\rho = 0.99$	61
34. Presentase Keragaman Komponen Utama dan Nilai eigen dengan $n = 300$ dan $\rho = 0.99$	62
35. Loadings SPCR Data $n = 300$ dan $\rho = 0.99$	63
36. Nilai Rata-rata Penduga Parameter, MMAD dan MSE dengan $n = 300$ dan $\rho = 0.99$	64
37. Perbandingan MMAD dan MSE Metode ENPCR dan SPCR dengan Seluruh Kelompok Data.....	66

DAFTAR GAMBAR

Gambar	Halaman
1. Perbandingan MSE ENPCR dan SPCR pada $\rho = 0.90$	67
2. Perbandingan MSE ENPCR dan SPCR pada $\rho = 0.99$	67

I. PENDAHULUAN

1.1 Latar Belakang dan Masalah

Analisis regresi adalah alat statistik yang digunakan untuk memperkirakan hubungan antara variabel respon dan variabel prediktor (Sial, 2021). Salah satu bentuk analisis regresi yang umum digunakan adalah regresi linear berganda, yang bertujuan untuk memahami hubungan antara satu variabel respon (terikat) dengan beberapa variabel prediktor (bebas) secara bersamaan. Namun, dalam dunia analisis data modern, salah satu tantangan yang sering muncul adalah masalah multikolinearitas. Menurut Gujarati & Porter (2009), salah satu syarat yang harus dipenuhi dalam analisis regresi linear berganda adalah tidak adanya multikolinearitas.

Multikolinearitas adalah fenomena ketika dua atau lebih variabel prediktor berkorelasi. Jika ini terjadi, kesalahan standar koefisien akan meningkat, menyebabkan hasil yang tidak akurat dalam analisis regresi (McClendon & McKee, 2002). Masalah ini dapat mengakibatkan nilai probabilitas dan interval kepercayaan koefisien regresi yang tidak dapat diandalkan (Kim, 2019). Kondisi ini umumnya terjadi pada dataset yang berisi banyak variabel, terutama dalam konteks fenomena sosial yang kompleks.

Untuk mengatasi masalah ini, metode tradisional seperti *Principal Component Regression* (PCR) sering digunakan (Sinaga, *et al.*, 2019). PCR bekerja dengan melakukan dekomposisi terhadap variabel prediktor menjadi beberapa komponen

utama yang tidak berkorelasi. Komponen tersebut kemudian digunakan dalam analisis regresi (Jolliffe, 2002).

Namun, PCR memiliki kelemahan, yaitu tidak mempertimbangkan relevansi komponen utama terhadap variabel respon dalam proses seleksinya. Hal ini dapat menyebabkan informasi penting untuk prediksi justru terabaikan (Cook & Forzani, 2008). Selain itu, PCR tidak memiliki mekanisme untuk memilih variabel prediktor yang paling signifikan, sehingga cenderung kurang efisien ketika diterapkan pada dataset berdimensi tinggi (Hucker & Wahl, 2022).

Elastic Net Principal Component Regression (ENPCR) dikembangkan untuk mengatasi kedua kelemahan tersebut. ENPCR, yang diperkenalkan oleh Hamada (2024), mengombinasikan regularisasi *Elastic Net*, yang menggabungkan penalti L_1 dan L_2 , dengan pendekatan PCR. Metode ini memanfaatkan PCA untuk mereduksi dimensi data, sekaligus memilih variabel prediktor yang relevan melalui proses regularisasi. Dengan mengatasi multikolinearitas dan mendorong sparsitas, ENPCR mampu meningkatkan akurasi prediksi dan efisiensi model, terutama pada dataset dengan dimensi tinggi dan korelasi antar variabel yang kuat.

Sementara itu, *Sparse Principal Component Regression* (SPCR), yang diperkenalkan oleh Kawano, *et al.* (2015), mengintegrasikan proses dekomposisi komponen utama dan analisis regresi dalam satu langkah. Tidak seperti PCR yang hanya mengandalkan komponen utama untuk menjelaskan variabilitas data, SPCR secara langsung mempertimbangkan hubungan antara komponen utama dan variabel respon. SPCR menghasilkan beban (*loadings*) komponen utama yang jarang (*sparse*), sehingga model yang dihasilkan lebih sederhana dan relevan secara prediktif. Metode ini sangat efektif dalam menangani multikolinearitas sekaligus mempertahankan informasi penting yang berkaitan dengan variabel respon.

Penelitian oleh Hamada (2024) menunjukkan performa ENPCR yang lebih baik dibandingkan dengan PCR dalam analisis data yang memiliki dimensi tinggi dan korelasi antar variabel yang kuat. Sementara itu, penelitian yang dilakukan oleh Kawano, *et al.* (2015) menunjukkan bahwa SPCR lebih efektif dibandingkan PCR.

Berdasarkan uraian di atas, penulis tertarik untuk membandingkan ENPCR dan SPCR dalam menangani multikolinearitas. Kriteria pembanding yang digunakan dalam penelitian ini adalah nilai *Median of Mean Absolute Deviations* (MMAD) dan *Mean Square Error* (MSE) untuk mengevaluasi performa masing-masing metode. Oleh karena itu, penulis mengangkat penelitian ini dengan judul "Perbandingan Metode *Elastic Net Principal Component Regression* (ENPCR) dan *Sparse Principal Component Regression* (SPCR) dalam Mengatasi Multikolinearitas."

1.2 Tujuan Penelitian

Tujuan dari penelitian ini adalah :

1. Menduga parameter regresi dengan menggunakan metode ENPCR dan SPCR.
2. Membandingkan nilai dugaan antara metode ENPCR dan SPCR.
3. Mengetahui performa ENPCR dan SPCR dalam menangani masalah multikolinearitas.
4. Menentukan metode yang paling efektif di antara ENPCR dan SPCR dalam mengatasi masalah multikolinearitas menggunakan kriteria MMAD dan MSE.

1.3 Manfaat Penelitian

Manfaat dari penelitian ini adalah untuk memperluas pengetahuan serta diharapkan dapat memberikan wawasan bagi para peneliti, mahasiswa, dan pembaca mengenai penggunaan metode ENPCR dan SPCR dalam mengatasi masalah multikolinearitas.

II. TINJAUAN PUSTAKA

2.1 Analisis Regresi

Analisis regresi adalah teknik independen untuk menyelidiki dan memodelkan hubungan antar variabel. Aplikasi regresi sangat banyak dan terjadi di hampir setiap bidang, termasuk teknik, ilmu fisika dan kimia, ekonomi, manajemen, ilmu hayati dan biologi, dan ilmu sosial (Montgomery, *et al.*, 2012). Regresi linear sederhana adalah analisis regresi yang hanya melibatkan satu variabel prediktor. Model regresi linear adalah sebagai berikut:

$$Y = \beta_0 + \beta_1 X + \varepsilon \quad (2.1)$$

dengan,

Y = variabel respon

X = variabel prediktor

β_0 = intersep (bilangan konstan)

β_1 = parameter regresi

ε = residual

2.2 Analisis Regresi Linear Berganda

Analisis regresi linier berganda merupakan analisis hubungan secara linier antara dua atau lebih variabel prediktor dengan satu variabel respon. Secara umum,

model regresi linear berganda dapat dinyatakan sebagai berikut (Draper & Smith, 1998).

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + \varepsilon \quad (2.2)$$

dengan,

β_j = koefisien regresi ke- j , di mana $j = 1, 2, \dots, p$

X_j = variabel prediktor ke- j

Jika terdapat sampel acak berukuran n , yaitu $(Y_i, X_{i1}, X_{i2}, \dots, X_{ip})$ di mana $i = 1, 2, \dots, n$ dari suatu populasi, maka model regresinya bisa dituliskan sebagai:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_p X_{ip} + \varepsilon_i \quad (2.3)$$

Dalam bentuk matriks, model regresi linear berganda dapat dituliskan sebagai berikut (Montgomery, *et al.*, 2012).

$$Y = X\beta + \varepsilon \quad (2.4)$$

dengan,

Y = vektor observasi dari variabel respon berukuran $n \times 1$

X = matriks variabel prediktor berukuran $n \times (p + 1)$

β = vektor koefisien regresi berukuran $(p + 1) \times 1$

Dalam regresi linear berganda yang akan diduga adalah β_0 dan β_j artinya ($\beta_1, \beta_2, \dots, \beta_p$). Persamaan linear untuk pendugaan garis regresi linear ditulis dalam bentuk sebagai berikut.

$$\hat{y}_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + \cdots + b_p x_{ip} \quad (2.5)$$

dengan,

$\hat{y}_i$ = nilai dugaan variabel respon

x_{ij} = nilai variabel prediktor ke- j pengamatan ke- i

- b_0 = titik potong garis regresi pada sumbu-y atau nilai dugaan $\hat{y}$ bila x sama dengan nol
- b_j = gradien garis regresi (perubahan nilai dugaan $\hat{y}$ per satuan perubahan nilai x) ke- j

2.3 Asumsi Analisis Regresi Linear

Menurut Drapper & Smith (1992) untuk dapat menarik kesimpulan yang tepat mengenai parameter β_0 dan β_j , penting untuk memenuhi asumsi-asumsi yang ada dalam model regresi. Asumsi-asumsi tersebut meliputi:

1. Nilai ε_j harus independen satu sama lain atau tidak berkorelasi. Asumsi pertama ini menyatakan bahwa variabel prediktor merupakan variabel acak dengan nilai tengah nol dan varians (σ^2) yang tidak diketahui. Dengan kata lain, $E(\varepsilon_j) = 0, Var(\varepsilon_j) = \sigma^2$ dan ε_j dan ε_i tidak berkorelasi, $j \neq i$, sehingga $cov(\varepsilon_j, \varepsilon_j) = 0$. Oleh karena itu, ε_j dianggap sebagai variabel acak yang normal dengan nilai tengah nol dan varians tertentu, atau dengan kata lain, $\varepsilon_j \sim N(0, \sigma^2)$.
2. Nilai tengah dari variabel respon Y adalah fungsi linier dari variabel prediktor X. Jika titik-titik dari nilai tengah yang berbeda dihubungkan, maka akan terbentuk garis lurus, yaitu $\mu(Y|X) = \beta_0 + \beta_j X_j$.
3. Homoskedasitas, yaitu ragam galat harus homogen, di mana nilai ragam antara galat ke- j dan galat ke- i harus sama. Secara matematis, ini dapat ditulis sebagai $Var(\varepsilon_j) = \sigma^2$.
4. Ragam galat harus terdistribusi normal dengan rata-rata nol dan varians tertentu. Asumsi keempat ini menyatakan bahwa untuk setiap kombinasi tetap dari variabel prediktor X, variabel respon Y berdistribusi normal, yang dikenal sebagai asumsi kenormalan. Dengan kata lain, $\varepsilon_j \sim N(0, \sigma^2)$.

2.4 Metode Kuadrat Terkecil (MKT)

Metode Kuadrat Terkecil (MKT) adalah teknik yang digunakan untuk mengestimasi parameter $\beta_0, \beta_1, \beta_2, \dots, \beta_p$ dengan meminimalkan jumlah kuadrat dari galat. Parameter $\beta_0, \beta_1, \beta_2, \dots, \beta_p$ tidak diketahui, sehingga perlu dicari nilai estimasinya (Montgomery, *et al.*, 2012).

Menurut Hastie, *et al.* (2008), pendekatan MKT digunakan untuk memperkirakan β dengan cara meminimalkan jumlah kuadrat galat (JKG). Untuk memperoleh nilai β , langkahnya adalah dengan meminimalkan bentuk kuadrat:

$$Q(\hat{\beta}) \sum_{i=1}^n \varepsilon_i^2 = \varepsilon_i^T \varepsilon_i = \sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^p x_{ij} \beta_j \right)^2 \quad (2.6)$$

Dalam bentuk matriks, persamaan tersebut menjadi

$$\boldsymbol{\varepsilon} = \mathbf{Y} - \mathbf{X}\boldsymbol{\beta} \quad (2.7)$$

Sehingga, perkalian matriks galat menjadi:

$$\begin{aligned} \boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon} &= (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \\ \boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon} &= \mathbf{Y}^T \mathbf{Y} - \mathbf{Y}^T \mathbf{X}\boldsymbol{\beta} - \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{Y} + \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{X}\boldsymbol{\beta} \\ &\quad (\text{karena } \mathbf{Y}^T \mathbf{X}\boldsymbol{\beta} = \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{Y}) \\ \boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon} &= \mathbf{Y}^T \mathbf{Y} - 2\mathbf{Y}^T \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{X}\boldsymbol{\beta} \end{aligned}$$

Selanjutnya, $\boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon}$ akan diturunkan secara parsial terhadap $\boldsymbol{\beta}$, kemudian disamadengankan dengan nol.

$$\begin{aligned} \frac{\partial \boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon}}{\partial \boldsymbol{\beta}} &= -2\mathbf{X}^T \mathbf{Y} + 2\mathbf{X}^T \mathbf{X}\boldsymbol{\beta} = \mathbf{0} \\ -\mathbf{X}^T \mathbf{Y} + \mathbf{X}^T \mathbf{X}\boldsymbol{\beta} &= \mathbf{0} \\ \mathbf{X}^T \mathbf{X}\boldsymbol{\beta} &= \mathbf{X}^T \mathbf{Y} \\ (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{X}\boldsymbol{\beta} &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y} \\ \boldsymbol{\beta} &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y} \end{aligned}$$

Sehingga didapat penduga untuk MKT sebagai berikut.

$$\hat{\beta} = (X^T X)^{-1} X^T Y \quad (2.8)$$

Sifat-sifat penduga metode kuadrat terkecil adalah sebagai berikut.

1. $\hat{\beta}$ linear

$\hat{\beta}$ linear jika $\hat{\beta}$ merupakan fungsi linear dari β .

$$\begin{aligned} \hat{\beta} &= (X^T X)^{-1} X^T Y \\ &= (X^T X)^{-1} X^T (X\beta + \varepsilon) \\ &= (X^T X)^{-1} X^T X\beta + (X^T X)^{-1} X^T \varepsilon \\ &= I\beta + (X^T X)^{-1} X^T \varepsilon \end{aligned}$$

2. $\hat{\beta}$ tak bias

$\hat{\beta}$ adalah penduga tak bias jika $E(\hat{\beta}) = \beta$

$$\begin{aligned} E(\hat{\beta}) &= E((X^T X)^{-1} X^T Y) \\ &= E((X^T X)^{-1} X^T (X\beta + \varepsilon)) \\ &= E((X^T X)^{-1} X^T X\beta + (X^T X)^{-1} X^T \varepsilon) \\ &= (X^T X)^{-1} X^T X\beta + (X^T X)^{-1} X^T E(\varepsilon) \\ &= (X^T X)^{-1} X^T X\beta \\ &= \beta \end{aligned}$$

Sehingga $\hat{\beta}$ merupakan penduga tak bias bagi β .

3. $\hat{\beta}$ memiliki variansi minimum

$$\begin{aligned} Var(\hat{\beta}) &= E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)^T] \\ &= E[((X^T X)^{-1} X^T Y - \beta)((X^T X)^{-1} X^T Y - \beta)^T] \\ &= E[((X^T X)^{-1} X^T (X\beta + \varepsilon) - \beta)((X^T X)^{-1} X^T (X\beta + \varepsilon) - \beta)^T] \\ &= E[((X^T X)^{-1} X^T X\beta + ((X^T X)^{-1} X^T \varepsilon) - \beta)((X^T X)^{-1} X^T X\beta + \\ &\quad ((X^T X)^{-1} X^T \varepsilon) - \beta)^T] \\ &= E[((X^T X)^{-1} X^T X\beta + ((X^T X)^{-1} X^T \varepsilon) - \beta)((X^T X)^{-1} X^T X\beta + \\ &\quad ((X^T X)^{-1} X^T \varepsilon) - \beta)^T] \end{aligned}$$

$$\begin{aligned}
&= E[(\boldsymbol{\beta} + (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\varepsilon}) - \boldsymbol{\beta})(\boldsymbol{\beta} + ((\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\varepsilon}) - \boldsymbol{\beta})^T] \\
&= E[((\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\varepsilon})(\boldsymbol{\beta} + ((\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\varepsilon})^T)] \\
&= E[(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon} \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1}] \\
&= E[(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T E(\boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon}) \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1}] \\
&= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} E(\boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon}) \\
&= (\mathbf{X}^T \mathbf{X})^{-1} \boldsymbol{\sigma}^2
\end{aligned}$$

$Var(\hat{\boldsymbol{\beta}}) = (\mathbf{X}^T \mathbf{X})^{-1} \boldsymbol{\sigma}^2$ merupakan merupakan variansi yang paling kecil di antara semua penduga linear yang tidak bias, sesuai dengan teorema Gauss-Markov.

Dengan demikian, apabila penduga kuadrat terkecil memenuhi sifat linear, tidak bias, dan memiliki variansi minimum disebut sebagai *Best Linear Unbiased Estimator* (BLUE).

2.5 Multikolinearitas

Istilah multikolinearitas pertama kali diperkenalkan oleh Ragnar Frisch pada tahun 1934. Awalnya, multikolinearitas merujuk pada adanya hubungan linear yang sempurna atau pasti antara beberapa atau seluruh variabel prediktor dalam model regresi berganda (Gujarati, 2009). Dalam hubungan ekonomi, multikolinearitas sering terjadi, sehingga perhatian lebih difokuskan pada dampaknya terhadap sampel daripada keberadaannya (Sumodiningrat, 1996).

Jika asumsi model regresi linear klasik terpenuhi, penduga MKT untuk koefisien regresi bersifat linear, tak bias, dan mempunyai variansi minimum atau *Best Linear Unbiased Estimator* (BLUE). Namun, meskipun multikolinearitas sangat tinggi, penduga MKT tetap memenuhi syarat sebagai BLUE, tetapi cenderung tidak stabil (Gujarati, 2009).

Menurut Montgomery *et al.* (2012), terdapat beberapa pendekatan untuk mendeteksi masalah multikolinearitas, di antaranya:

1. Memeriksa korelasi antar variabel prediktor

Masalah multikolinearitas dapat diindikasikan melalui tingginya korelasi antar variabel prediktor. Jika nilai koefisien korelasi antar variabel prediktor lebih dari 0,5, kemungkinan besar terdapat multikolinearitas atau kolinearitas ganda di antara variabel tersebut.

2. Menggunakan *Variance Inflation Factor* (VIF)

VIF merupakan salah satu metode alternatif untuk mengidentifikasi keberadaan multikolinearitas. Nilai variansi meningkat seiring dengan pengaruh σ^2 dan VIF itu sendiri. Jika VIF bernilai lebih dari 10, hal ini menunjukkan adanya multikolinearitas yang signifikan antara variabel prediktor. Adapun rumus yang digunakan untuk menghitung VIF adalah:

$$VIF = \frac{1}{1 - R_i^2} \quad (2.9)$$

R_j^2 merupakan koefisien determinasi ke- j ($j = 1, 2, \dots, p$) yang diperoleh dari variabel prediktor X_j yang diestimasi dengan menggunakan variabel prediktor lainnya.

2.6 Standarisasi Variabel

Proses standarisasi variabel bertujuan untuk mengubah data sehingga variabel memiliki satuan yang seragam atau perbandingan yang lebih mudah dilakukan antar variabel. Modifikasi sederhana dari proses ini adalah transformasi korelasi. Proses ini melibatkan dua tahap utama: pertama, mengubah data dengan mengurangi nilai rata-rata pengamatan untuk setiap variabel, dan kedua, mengubah data ke dalam satuan deviasi standar. (Kutner, *et al.*, 2005).

Dalam hal ini yang akan distandarisasi adalah model regresi linear berganda pada persamaan (2.3). Berikut merupakan standarisasi variabel respon Y dan variabel prediktor $X_1, X_2, \dots, X_p$.

$$y_i^* = \frac{y_i - \bar{Y}}{S_Y} \text{ dengan } S_Y = \sqrt{\frac{\sum_{i=1}^n (y_i - \bar{Y})^2}{n-1}} \quad (2.10)$$

$$x_{ij}^* = \frac{x_{ij} - \bar{X}_j}{S_{X_j}} \text{ dengan } S_{X_j} = \sqrt{\frac{\sum_{i=1}^n (x_{ij} - \bar{X}_j)^2}{n-1}} \quad (2.11)$$

dengan,

- $\bar{Y}$ = rata-rata dari Y
- $\bar{X}_j$ = rata-rata dari pengamatan X_j
- S_Y = standar deviasi dari Y
- S_{X_j} = standar deviasi dari X

Model regresi berganda terstandarisasi merupakan hasil transformasi dari model regresi berganda (dijelaskan sebagai transformasi korelasi).

$$y_i^* = \beta_1^* x_{i1}^* + \beta_2^* x_{i2}^* + \dots + \beta_p^* x_{ip}^* + \varepsilon_i^* \quad (2.12)$$

Model tersebut dikenal sebagai model regresi terstandarisasi (*standardized regression model*). Antara parameter $\beta_1^*, \beta_2^*, \dots, \beta_p^*$ dalam model regresi terstandarisasi dan parameter $\beta_1, \beta_2, \dots, \beta_p$ dalam model regresi linear berganda, terdapat hubungan linear. Hubungan antara parameter dari kedua model tersebut dijelaskan seperti berikut (Kutner, *et al.*, 2005).

$$\beta_j = \left(\frac{S_Y}{S_{X_j}} \right) \beta_j^*, \quad j = 1, 2, \dots, p \quad (2.13)$$

$$\beta_0 = \bar{Y} - \beta_1 \bar{X}_1 - \beta_2 \bar{X}_2 - \dots - \beta_p \bar{X}_p \quad (2.14)$$

$$\beta_0 = \bar{Y} - \sum_{j=1}^p \beta_j \bar{X}_j \quad (2.15)$$

Prosedur ini dikenal sebagai prosedur penyesuaian skala (*scaling*). Dari persamaan (2.3), bentuknya dapat diubah menjadi:

$$\begin{aligned}
 y_i &= \beta_0 + \beta_1(x_{i1} - \bar{X}_1) + \beta_1\bar{X}_1 + \beta_2(x_{i2} - \bar{X}_2) + \beta_2\bar{X}_2 + \dots + \\
 &\quad \beta_p(x_{ip} - \bar{X}_p) + \beta_p\bar{X}_p + \varepsilon_i \\
 y_i &= (\beta_0 + \beta_1\bar{X}_1 + \dots + \beta_p\bar{X}_p) + \beta_1(x_{i1} - \bar{X}_1) + \dots + \beta_p(x_{ip} - \bar{X}_p) + \varepsilon_i \quad (2.16)
 \end{aligned}$$

Berdasarkan persamaan (2.15) maka:

$$\bar{Y} = \beta_0 + \beta_1\bar{X}_1 + \beta_2\bar{X}_2 + \dots + \beta_p\bar{X}_p$$

sehingga, jika $\bar{Y}$ dikurangkan dari Y_i , diperoleh:

$$\begin{aligned}
 y_i - \bar{Y} &= (\beta_0 + \beta_1\bar{X}_1 + \dots + \beta_p\bar{X}_p) + (\beta_1(x_{i1} - \bar{X}_1) + \dots + \\
 &\quad \beta_p(x_{ip} - \bar{X}_p) + \varepsilon_i) - (\beta_0 + \beta_1\bar{X}_1 + \beta_2\bar{X}_2 + \dots + \beta_p\bar{X}_p) \\
 y_i - \bar{Y} &= \beta_1(x_{i1} - \bar{X}_1) + \beta_2(x_{i2} - \bar{X}_2) + \dots + \beta_p(x_{ip} - \bar{X}_p) + \varepsilon_i \quad (2.17)
 \end{aligned}$$

dengan substitusi:

$$\begin{aligned}
 y_i &= y_i - \bar{Y} \\
 x_{i1} &= x_{i1} - \bar{X} \\
 x_{i2} &= x_{i2} - \bar{X} \\
 x_{ip} &= x_{ip} - \bar{X}
 \end{aligned}$$

maka didapat model baru yaitu:

$$y_i = \beta_1x_{i1} + \beta_2x_{i2} + \dots + \beta_px_{ip} + \varepsilon_i \quad (2.18)$$

Prosedur ini disebut prosedur penyesuaian rata-rata (*centering*), di mana β_0 (*intercept*) dihilangkan, sehingga penyusunan model regresi menjadi lebih sederhana. Keseluruhan proses ini disebut prosedur penyesuaian rata-rata dan skala (*centering & scaling*).

2.7 Principal Component Analysis (PCA)

Principal Component Analysis (PCA) atau Analisis Komponen Utama (AKU) merupakan teknik statistik yang secara linear mengubah sekumpulan variabel asli menjadi sejumlah variabel yang lebih kecil dan tidak berkorelasi, yang tetap merepresentasikan informasi dari variabel-variabel asli (Radiarta, *et al.*, 2013). Menurut Pratiwi & Roosyanti (2019) metode PCA bertujuan untuk menyederhanakan data yang besar menjadi dimensi yang lebih sederhana dengan mereduksi variabel tanpa menghilangkan informasi dari variabel asli.

Menurut Jolliffe (2002) komponen utama adalah kombinasi linier dari variabel prediktor asli $(X_1, X_2, \dots, X_p)$, yang dapat dituliskan dalam bentuk matriks sebagai berikut.

$$\mathbf{Z} = \mathbf{XV} \quad (2.19)$$

dengan,

- $\mathbf{Z}$ = matriks komponen utama berukuran $n \times p$, di mana n adalah jumlah observasi dan k adalah jumlah komponen utama yang dipilih
- $\mathbf{X}$ = matriks variabel prediktor berukuran $n \times p$
- $\mathbf{V}$ = matriks *loadings* atau koefisien dengan ukuran $p \times p$ yang menggambarkan kontribusi masing-masing variabel prediktor terhadap komponen utama

Menurut Jolliffe (2002), $\mathbf{V}$ adalah matriks ortogonal yang tersusun dari vektor *eigen* yang sesuai dengan nilai *eigen* dari matriks $\mathbf{X}^T \mathbf{X}$. Ini berarti bahwa $\mathbf{V}$ memenuhi sifat:

$$\mathbf{VV}^T = \mathbf{V}^T \mathbf{V} = \mathbf{I}_p \quad (2.20)$$

Karena sifat ortogonal ini, $\mathbf{X}$ dapat ditulis ulang dalam bentuk komponen utama $\mathbf{Z}$, yaitu:

$$\begin{aligned} \mathbf{XV} &= \mathbf{Z} \\ \mathbf{XVV}^T &= \mathbf{ZV}^T \\ \mathbf{X} &= \mathbf{ZV}^T \end{aligned} \quad (2.21)$$

Hal ini menunjukkan bahwa variabel prediktor $\mathbf{X}$ dapat direpresentasikan sebagai kombinasi linier dari komponen utama $\mathbf{Z}$, yang memberikan pandangan komprehensif mengenai informasi dari variabel asli melalui komponen utama (Lee, *et al.*, 2015).

Menurut Johnson & Wichern (2007) komponen utama tidak berkorelasi dan memiliki varians yang sama dengan nilai *eigen* dari matriks kovarians Σ dari $\mathbf{X}$. Oleh karena itu, nilai *eigen* terbesar menunjukkan varians terbesar. Terdapat tiga kriteria dalam pemilihan komponen utama yang digunakan, yaitu:

- 1) Memilih nilai *eigen* yang lebih besar dari satu di mana nilai *eigen* yang mendekati nol dianggap tidak memiliki pengaruh signifikan.
- 2) Memperhatikan sudut pada *scree plot*, yaitu grafik yang menggambarkan nilai *eigen* dan menunjukkan perubahan signifikan dalam nilainya.
- 3) Proporsi variansi yang cukup untuk mewakili total variansi kumulatif minimal 80 persen.

Transformasi ke komponen utama ini membantu menghilangkan multikolinearitas antar variabel tanpa harus mengurangi jumlah variabel asal (Rougoor, *et al.*, 2000). Komponen utama pertama, Z_1 , merupakan kombinasi linear dari variabel asal yang menjelaskan varians terbesar dalam data (Gaspersz, 1991). Secara umum, komponen utama ke- j dapat ditulis sebagai:

$$Z_j = v_{1j}X_1 + v_{2j}X_2 + \cdots + v_{pj}X_p = \mathbf{v}_j^T \mathbf{X} \quad (2.22)$$

Setelah beberapa komponen utama yang bebas dari multikolinearitas diperoleh, komponen-komponen tersebut dapat digunakan sebagai variabel prediktor baru dalam analisis regresi untuk memahami pengaruhnya terhadap variabel respons (Y). PCA memungkinkan variabel-variabel ini untuk lebih mudah diinterpretasikan dan dapat menghasilkan estimasi yang lebih stabil dalam konteks regresi.

2.8 *Principal Component Regression (PCR)*

Menurut Jolliffe (2002) *Principal Component Regression (PCR)* adalah metode regresi yang mengubah model regresi standar dengan menggunakan komponen utama sebagai variabel respon. Tujuan utama PCR adalah untuk mengatasi multikolinearitas antara variabel prediktor dengan mengubah variabel-variabel asli menjadi sekumpulan komponen utama yang tidak berkorelasi. Dengan mereduksi dimensi variabel prediktor tanpa kehilangan informasi yang signifikan, PCR menyederhanakan analisis dan menghasilkan model yang lebih stabil.

Menurut Montgomery, *et al.* (2012), bentuk umum dari model PCR dapat dinyatakan sebagai berikut.

$$Y = Z\gamma + \varepsilon \quad (2.23)$$

di mana, $Z = XV$ dan $\gamma = V^T\beta$ dengan,

Y = vektor variabel respon

X = matriks yang elemen-elemennya dikurang dengan rataannya yang mensyaratkan rataan 0 dan standar deviasi 1 (standarisasi)

Z = matriks berukuran $n \times p$ yang elemennya terdapat komponen utama

γ = vektor koefisien regresi yang terkait dengan komponen utama Z

ε = vektor galat berukuran $n \times 1$

PCR memanfaatkan estimasi kuadrat terkecil, tetapi menggunakan skor komponen utama sebagai variabel prediktor. Secara umum, estimasi regresi dalam PCR dapat diformulasikan sebagai berikut:

$$\hat{\gamma} = \arg \min_{\gamma} \|y - Z\gamma\|_2^2 = \arg \min_{\gamma} \left\| y - \sum_{i=1}^n \gamma_i \tilde{z}_i \right\|_2^2 \quad (2.24)$$

Sehingga

$$Q(\hat{\gamma}) = \sum_{i=1}^n (y_i - \hat{\gamma}_0 - \hat{\gamma}_1 z_{1i} - \hat{\gamma}_2 z_{2i} - \dots - \hat{\gamma}_p z_{pi})^2$$

Langkah berikutnya yaitu meminimumkan fungsi $Q(\hat{\gamma})$ dan menyamakannya dengan nol. Fungsi objektif ini dapat disusun dalam bentuk matriks sebagai berikut.

$$\begin{aligned} [Z^T Z] \hat{\gamma} &= Z^T y \\ [Z^T Z]^{-1} \hat{\gamma} &= [Z^T Z]^{-1} Z^T y \\ \hat{\gamma} &= [Z^T Z]^{-1} Z^T y \end{aligned} \quad (2.25)$$

$$\hat{\beta} = V \hat{\gamma} \quad (2.26)$$

Dengan variabel-variabel yang sudah distandarisasi, $X^T X$ menjadi matriks korelasi antar prediktor, sedangkan $X^T Y$ menjadi vektor korelasi antara prediktor dan respon. Misalkan $V = [V_1, V_2, \dots, V_p]$ adalah matriks vektor *eigen* dari $X^T X$, dan $\tau_1, \dots, \tau_p$ adalah nilai *eigen* yang bersesuaian.

Misalkan $Z = [Z_1, \dots, Z_p] = XV$. Sehingga $Z_j = XV_j$ adalah komponen utama ke- j dari X . Beberapa hal penting dari regresi komponen utama yaitu:

1. $V^T V = V V^T = I_p$, di mana matriks V ortonormal
2. $Z^T Z = \Lambda$ di mana $\Lambda = \text{diag}(\tau_1, \dots, \tau_p)$, Z orthogonal dan $|Z_j| = \sqrt{\tau_j}$
3. $X = Z V^T$ dan $X_i = \sum_{k=1}^p v_{jk} Z_k$.

Dengan asumsi bahwa $ZV^T = \delta$, maka model PCR dapat dituliskan sebagai berikut (Perez, 2017).

$$\begin{aligned}
 Y &= X\beta + \varepsilon \\
 Y &= XVV^T\beta + \varepsilon \\
 Y &= ZV^T\beta + \varepsilon \\
 Y &= \delta\beta + \varepsilon
 \end{aligned} \tag{2.27}$$

Di sini, δ merepresentasikan kombinasi linear dari komponen utama yang terkait langsung dengan variabel prediktor asli, dan vektor β adalah parameter regresi yang ingin diestimasi.

Estimasi parameter β dalam PCR dilakukan dengan meminimalkan jumlah kuadrat galat antara Y dan $\delta\beta$. Fungsi objektif yang diminimalkan adalah:

$$\hat{\beta} = \arg \min_{\beta} \|Y - \delta\beta\|_2^2 \tag{2.28}$$

Dengan mengambil turunan fungsi ini terhadap β dan menyetarakannya dengan nol, diperoleh estimasi β sebagai berikut (Jolliffe, 2002).

$$\hat{\beta} = (\delta^T \delta)^{-1} \delta^T Y \tag{2.29}$$

Lakukan substitusi sehingga hasil estimasi beta dapat diturunkan sebagai berikut.

$$\begin{aligned}
 \hat{\beta} &= (\delta^T \delta)^{-1} \delta^T Y \\
 \hat{\beta} &= ((ZV^T)^T (ZV^T))^{-1} (ZV^T)^T Y \\
 \hat{\beta} &= (VZ^T ZV^T)^{-1} VZ^T Y \\
 \hat{\beta} &= V(Z^T Z)^{-1} V^T VZ^T Y \\
 \hat{\beta} &= V(Z^T Z)^{-1} Z^T Y \\
 \hat{\beta} &= V\hat{\gamma}
 \end{aligned} \tag{2.30}$$

Jika semua komponen utama digunakan, estimasi regresi komponen utama sama dengan regresi kuadrat terkecil. Namun, dalam praktiknya, hanya sebagian komponen utama yang dipilih, $\mathbf{Z}_{(s)} = [Z_1, \dots, Z_k]$ sehingga estimasi menjadi:

$$\hat{\mathbf{y}} = (\mathbf{Z}_{(s)}^T \mathbf{Z}_{(s)})^{-1} \mathbf{Z}_{(s)}^T \mathbf{Y} = \mathbf{A}_{(s)}^{-1} \mathbf{Z}_{(s)}^T \mathbf{Y} \quad (2.31)$$

2.9 Elastic Net Principal Component Regression (ENPCR)

Menurut Hamada (2024), *Elastic Net Principal Component Regression* (ENPCR) merupakan pendekatan yang mengintegrasikan keuntungan dari PCR dengan strategi regularisasi *Elastic Net*. PCR sering digunakan untuk mengatasi masalah multikolinearitas dengan cara mengubah variabel-variabel yang berkorelasi menjadi sekumpulan komponen utama yang tidak berkorelasi. Namun, PCR tidak selalu efektif dalam melakukan seleksi variabel penting, terutama dalam konteks data berdimensi tinggi.

Elastic Net adalah metode regularisasi yang menggabungkan keuntungan dari *Lasso* (yang menerapkan penalti L_1 untuk mempromosikan sparsitas) dan *Ridge* (yang menerapkan penalti L_2 untuk menstabilkan estimasi), sehingga meningkatkan performa prediksi sekaligus mengatasi multikolinearitas. Dengan cara ini, *Elastic Net* dapat secara efektif memilih variabel penting meskipun terdapat korelasi tinggi antar prediktor (Farhadi, *et al.*, 2019).

Persamaan yang digunakan dalam ENPCR adalah sebagai berikut (Hamada, 2024).

$$\hat{\boldsymbol{\beta}} = \arg \min_{\boldsymbol{\beta}} \|\mathbf{Y} - \boldsymbol{\delta} \boldsymbol{\beta}\|_2^2 + \lambda_1 \|\boldsymbol{\beta}\|_1 + \lambda_2 \|\boldsymbol{\beta}\|_2^2 \quad (2.32)$$

di mana $\boldsymbol{\delta} = \mathbf{ZV}^T$, dengan,

$\mathbf{Y}$ = vektor variabel respon

- δ = kombinasi linear dari komponen utama $\mathbf{Z}$ yang terkait langsung dengan variabel prediktor asli $\mathbf{X}$
- $\mathbf{Z}$ = matriks skor komponen utama
- $\mathbf{V}$ = matriks *eigen* vektor dari matriks korelasi $\mathbf{X}^T \mathbf{X}$
- $\hat{\beta}$ = penduga parameter regresi
- λ_1 = parameter penalti untuk norma L_1 dengan nilai 0 sampai 1
- λ_2 = parameter penalti untuk norma L_2 dengan nilai 0 sampai 1

Parameter λ_1 mendorong sparsitas dengan mengatur koefisien menjadi nol, dan λ_2 menstabilkan estimasi untuk mengurangi multikolinearitas. ENPCR efektif ketika jumlah prediktor lebih banyak dari observasi atau ada korelasi tinggi antar prediktor, dengan memungkinkan seleksi variabel yang relevan dan menghasilkan model yang stabil dan mudah diinterpretasikan.

Parameter λ_1 dan λ_2 dioptimalkan menggunakan *K-fold cross-validation*, di mana dataset dibagi menjadi K subset. Kriteria optimasi dihitung dengan:

$$CV = \frac{1}{K} \sum_{i=1}^k (y_i - \hat{y}_{-i(k)})^2$$

di mana $\hat{y}_{-i(k)}$ adalah estimasi y_i saat *fold* ke- k tidak digunakan dalam pelatihan model.

2.10 Sparse Principal Component Regression (SPCR)

Sparse Principal Component Regression (SPCR) merupakan metode regresi yang dikembangkan oleh Kawano, *et al.* (2015) untuk menangani masalah multikolinearitas dan pemilihan variabel secara simultan. Metode ini menggabungkan prinsip regresi linier dengan teknik reduksi dimensionalitas berbasis komponen utama yang jarang (*sparse principal components*). Dalam

SPCR, hanya variabel dengan kontribusi signifikan yang dipertahankan, sehingga model yang dihasilkan menjadi lebih sederhana dan interpretatif.

Pendekatan SPCR menggunakan komponen utama $\mathbf{M}^T \mathbf{X}$ didasarkan pada formulasi berikut (Kawano, *et al.*, 2015).

$$\min_{\mathbf{A}, \mathbf{M}, \gamma_0, \boldsymbol{\gamma}} \left\{ (1-w) \sum_{i=1}^n (y_i - \gamma_0 - \boldsymbol{\gamma}^T \mathbf{M}^T \mathbf{x}_i)^2 + w \sum_{i=1}^n \|\mathbf{x}_i - \mathbf{A} \mathbf{M}^T \mathbf{x}_i\|^2 + \lambda_m (1-\xi) \sum_{h=1}^k \|\mathbf{m}_h\|_1 + \lambda_h \xi \sum_{h=1}^k \|\mathbf{m}_h\|^2 + \lambda_\gamma \|\boldsymbol{\gamma}\|_1 \right\} \quad (2.33)$$

dengan syarat $\mathbf{A}^T \mathbf{A} = \mathbf{I}_k$, di mana:

$\mathbf{M}$ = matriks *loadings* berukuran $p \times k$, $\mathbf{M} = (\mathbf{m}_1, \dots, \mathbf{m}_k)$

$\mathbf{A}$ = matriks $p \times k$, $\mathbf{A} = (\alpha_1, \dots, \alpha_k)$.

γ_0 = *intercept*

$\boldsymbol{\gamma}$ = vektor koefisien, $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \dots, \gamma_k)^T$

Persamaan (2.33) dalam SPCR menggabungkan fungsi loss regresi linear dan analisis komponen utama (PCA) ke dalam satu formulasi objektif. Fungsi loss regresi linear bertujuan untuk meminimalkan perbedaan antara nilai aktual dan prediksi, dengan menghubungkan variabel respons (Y) ke komponen utama yang dihasilkan, sehingga meningkatkan akurasi prediksi. Di sisi lain, PCA memungkinkan ekstraksi komponen utama dari data prediktor (X) untuk menangkap informasi penting, sambil memperkenalkan sparsitas pada matriks *loadings* ($\mathbf{M}$).

Sparsitas pada model dicapai melalui regularisasi dengan norma L_1 dan L_2 , yang menekan beberapa parameter menjadi nol guna menyederhanakan model dan meningkatkan interpretabilitas. Parameter pengatur w dalam formulasi ini mengatur keseimbangan antara fokus pada akurasi prediksi dan pengurangan dimensionalitas data, sementara parameter ξ mengontrol proporsi penggunaan

norma L_1 dan L_2 untuk menjaga stabilitas model. Kombinasi elemen-elemen ini menjadikan SPCR efektif dalam menangani data berdimensi tinggi dengan hasil yang mudah diinterpretasi.

Persamaan (2.33) dapat ditulis ulang sebagai berikut.

$$\min_{\mathbf{A}, \mathbf{M}, \gamma_0, \boldsymbol{\gamma}} \left[(1-w) \sum_{i=1}^n \left\{ y_i - \gamma_0 - \sum_{h=1}^k \gamma_h \left(\sum_{j=1}^p m_{hj} x_{ij} \right) \right\}^2 + w \sum_{h=1}^k \sum_{i=1}^n \left(y_{hi}^* - \sum_{h=1}^k m_{hj} x_{ih} \right) + \lambda_m (1-\xi) \sum_{h=1}^k \sum_{j=1}^p |m_{hj}| + \lambda_h \xi \sum_{h=1}^k \sum_{j=1}^p m_{hj}^2 + \lambda_\gamma \sum_{h=1}^k |\gamma_h| \right]$$

dengan syarat $\mathbf{A}^T \mathbf{A} = \mathbf{I}_k$ di mana y_{hi}^* adalah elemen ke- i dari vektor $\mathbf{Xm}_h$.

Estimasi parameter $\mathbf{A}$ dilakukan dengan menggunakan algoritma oleh Zou, *et al.* (2006). Sementara itu, parameter $\mathbf{M}$ dan $\boldsymbol{\gamma}$ diestimasi melalui algoritma *coordinate descent* (Friedman, *et al.*, 2010), yang memperbarui satu parameter pada satu waktu sambil membekukan yang lain, hingga fungsi objektif diminimalkan, mengingat masalah optimisasi tersebut mencakup term regularisasi L_1 . Algoritma detailnya adalah sebagai berikut.

Pembaruan m_{jh} dengan γ_0, γ_h dan $\mathbf{A}$ diberikan. Pembaruan secara koordinat untuk m_{jh} memiliki bentuk:

$$\hat{m}_{j'h'} \leftarrow \frac{S \left(\sum_{i=1}^n x_{ij'} + \{ (1-w) Y_i \gamma_{h'} + Y_{h'i}^* w \}, \frac{\lambda_h (1-\xi)}{2} \right)}{\{ (1-w) \gamma_{h'}^2 + w \} \sum_{i=1}^n x_{ij'}^2 + \lambda_h \xi} \quad (2.34)$$

dengan $(j' = 1, \dots, p; h' = 1, \dots, k)$, di mana

$$Y_i = y_i - \gamma_0 - \sum_{h=1}^k \sum_{j \neq j'} \gamma_h m_{jh} x_{ij} - \sum_{h \neq h'} \gamma_h m_{j'h} x_{ij'}$$

$$Y_{h'i}^* = y_{h'i}^* - \sum_{l \neq l'} m_{jl'} x_{il'}$$

dan $S(u, \eta)$ adalah operator *soft-thresholding* dengan nilai:

$$\text{sign}(u)(|u| - \eta)_+ = \begin{cases} u - \eta & (u > 0 \text{ dan } \eta < |u|) \\ u + \eta & (u < 0 \text{ dan } \eta < |u|) \\ 0 & (\eta \geq |u|) \end{cases}$$

Pembaruan γ_h dengan γ_0, m_{hj} dan $\mathbf{A}$ diberikan. Pembaruan untuk γ_h ditulis sebagai:

$$\hat{\gamma}_{h'} \leftarrow \frac{S\left((1-w) \sum_{i=1}^n y_i^{**} X_{ih'}^*, \frac{\lambda_\gamma}{2}\right)}{(1-w) \sum_{i=1}^n x_{ih'}^2}, \quad (h' = 1, \dots, k) \quad (2.35)$$

dengan $x_{ih}^* = \mathbf{m}_h^T \mathbf{x}_i$ dan $y_i^{**} = y_i - \gamma_0 - \sum_{h \neq h'} \gamma_h x_{ih}^*$.

Pembaruan $\mathbf{A}$ dengan γ_0, m_{hj} dan γ_h diberikan. Estimasi untuk $\mathbf{A}$ diperoleh dengan melakukan dekomposisi nilai singular (SVD) pada $(\mathbf{X}^T \mathbf{X})\mathbf{M}$.

$$\hat{\mathbf{A}} = \mathbf{U}\mathbf{V}^T \quad (2.36)$$

di mana $(\mathbf{X}^T \mathbf{X})\mathbf{M} = \mathbf{U}\mathbf{D}\mathbf{V}^T$.

Pembaruan $\hat{\gamma}_0$ dengan m_{hj}, γ_h , dan $\mathbf{A}$ diberikan. Estimasi $\hat{\gamma}_0$ diperoleh dengan menghitung rata-rata dari selisih antara y_i dan kombinasi linear dari komponen lain, dengan rumus sebagai berikut.

$$\hat{\gamma}_0 = \frac{1}{n} \sum_{i=1}^n \left\{ y_i - \sum_{h=1}^k \hat{\gamma}_h \left(\sum_{j=1}^p \hat{m}_{jh} x_{ij} \right) \right\} \quad (2.37)$$

Langkah – langkah ini diulangi hingga konvergen.

Parameter *tuning* w dan ξ memainkan peran penting dalam akurasi prediksi.

Meskipun nilai w yang lebih kecil dapat menghasilkan prediksi yang lebih baik, model estimasi seringkali menjadi tidak stabil akibat fleksibilitas matriks $\mathbf{M}$.

Sementara itu, parameter *tuning* ξ berfungsi sebagai faktor *trade-off* antara penalti L_1 dan L_2 pada matriks $\mathbf{M}$. Nilai w dan ξ berada di antara 0 dan 1. Parameter λ_h dan λ_γ dioptimalkan melalui metode *K-fold cross-validation*. Dalam proses ini, dataset asli dibagi menjadi K subset $(y^{(1)}, X^{(1)}, \dots, y^{(K)}, X^{(K)})$. Kriteria untuk *cross-validation* dinyatakan sebagai:

$$CV = \frac{1}{K} \sum_{l=1}^K \left\| y^{(l)} - \hat{\gamma}_0^{(-l)} \mathbf{1}^{(l)} - \mathbf{X}^{(l)} \hat{\mathbf{M}}^{(-l)} \hat{\boldsymbol{\gamma}}^{(-l)} \right\|^2 \quad (2.38)$$

di mana $\mathbf{1}^{(l)}$ adalah vektor yang seluruh elemennya bernilai satu, dan $\hat{\gamma}_0^{(-l)}$, $\hat{\mathbf{M}}^{(-l)}$ dan $\hat{\boldsymbol{\gamma}}^{(-l)}$ merupakan estimasi yang diperoleh dengan mengeluarkan subset ke- l dari analisis.

Estimasi $\boldsymbol{\beta}$ mengikuti persamaan (2.26) yang menghasilkan estimasi parameter dalam bentuk

$$\hat{\boldsymbol{\beta}} = \mathbf{M} \hat{\boldsymbol{\gamma}} \quad (2.39)$$

2.11 Ketepatan Model

Ketepatan model dapat ditinjau dari nilai *Median of Mean Absolute Deviations* (MMAD) (Hamada, 2024). Selain itu, menurut Montgomery, *et al.* (2012) pemilihan model yang paling optimal dapat dilakukan dengan meninjau nilai *Mean Squared Error* (MSE). MSE merupakan salah satu pendekatan yang sederhana dan sering digunakan untuk mengukur tingkat kesalahan prediksi. Perhitungan MSE dilakukan dengan cara mengkuadratkan selisih antara nilai prediksi dan nilai sebenarnya. Semakin rendah nilai MMAD dan MSE, semakin tinggi akurasi prediksi tersebut. Efisiensi dari metode untuk menangani multikolinearitas ini akan dievaluasi berdasarkan nilai MMAD dan MSE dari hasil estimasi parameter $\boldsymbol{\beta}$, yang didefinisikan sebagai berikut :

$$MMAD = median \left(\frac{1}{p} \sum_{j=1}^p |\boldsymbol{\delta}^T \widehat{\boldsymbol{\beta}}_g - \boldsymbol{\delta}^T \boldsymbol{\beta}| \right) \quad (2.40)$$

$$MSE = \frac{1}{G} \sum_{g=1}^G \|\widehat{\boldsymbol{\beta}}_g - \boldsymbol{\beta}\|^2 \quad (2.41)$$

dengan,

$\boldsymbol{\beta}$ = parameter regresi untuk variabel prediktor

$\widehat{\boldsymbol{\beta}}_g$ = penduga parameter regresi pada ulangan ke- g

G = banyaknya ulangan.

III. METODOLOGI PENELITIAN

3.1 Tempat dan Waktu Penelitian

Penelitian ini dilaksanakan pada tahun ajaran 2024/2025 bertempat di Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.

3.2 Data Penelitian

Data yang digunakan dalam penelitian ini adalah data simulasi. Untuk data simulasi ini yakni data menggunakan model dense (padat) mengikuti rumus:

$$y_i = x_{1i} + x_{2i} + x_{3i} + x_{4i} + x_{5i} + x_{6i} + x_{7i} + x_{8i} + x_{9i} + \varepsilon_i \quad (3.1)$$
$$i = 1, 2, \dots, n.$$

Di sini, variabel prediktor sebanyak $p = 9$ dibangkitkan dari distribusi normal multivariat ($X \sim N_9(0, \Sigma_9)$) yang memiliki rata-rata nol (vektor nol) dan matriks kovarian Σ_p dengan $\beta_0 = 0$ dan $\beta_1 = \beta_2 = \dots = \beta_9 = 1$.

Dalam studi simulasi ini, dua jenis koefisien korelasi digunakan antara variabel prediktor, yaitu $\rho_1 = 0.90$ dan $\rho_2 = 0.99$ serta tiga ukuran sampel yang berbeda diuji yaitu $n = 100, 200$ dan 300 yang diulang sebanyak 100 kali.

3.3 Metode Penelitian

Penelitian ini dilakukan melalui studi pustaka dengan mempelajari buku-buku referensi serta karya ilmiah dalam bentuk jurnal. Dalam penelitian ini, penulis menggunakan perangkat lunak R-Studio versi 4.4.2 untuk mempercepat proses perhitungan dan menghasilkan hasil yang akurat.

Adapun langkah-langkah yang dilakukan pada penelitian ini adalah sebagai berikut.

1. Membangkitkan data.
2. Melakukan standarisasi variabel pada data.
3. Membuat matriks korelasi dari data yang sudah ditransformasi dan distandarisasi.
4. Mencari nilai *eigen* dan vektor *eigen* dari matriks korelasi.
5. Menghitung skor komponen utama $\mathbf{Z}$.
6. Memilih komponen utama yang memiliki nilai *eigen* lebih dari 1.
7. Melakukan analisis dengan *Elastic Net Principal Component Regression* (ENPCR) dengan langkah -langkah sebagai berikut.
 - a. Melakukan *Cross Validation* (CV) dengan 5-fold untuk mendapatkan nilai optimal λ_1 dan λ_2 .
 - b. Melakukan estimasi β ENPCR berdasarkan komponen utama yang terpilih dan λ_1 dan λ_2 hasil CV.
 - c. Mentransformasi estimasi β ENPCR ke dalam bentuk variabel asli.
 - d. Mencatat estimasi β metode ENPCR.
8. Melakukan analisis dengan *Sparse Principal Component Regression* (SPCR) dengan langkah -langkah sebagai berikut.
 - a. Melakukan *Cross Validation* (CV) dengan 5-fold untuk mendapatkan nilai optimal ξ, w, λ_m dan λ_γ .
 - b. Melakukan estimasi $\gamma, \mathbf{A}$ dan $\mathbf{M}$ dengan menggunakan nilai ξ, w, λ_m dan λ_γ hasil CV.
 - c. Melakukan estimasi β SPCR berdasarkan hasil estimasi γ dan $\mathbf{M}$.

- d. Mentransformasi estimasi β SPCR ke dalam bentuk variabel asli.
 - e. Mencatat estimasi β metode SPCR.
9. Ulangi langkah 1-8 sebanyak 100 kali.
 10. Menghitung nilai MMAD dan MSE β ENPCR dan SPCR.
 11. Membandingkan nilai MMAD dan MSE metode ENPCR dan SPCR untuk menentukan metode terbaik.

V. KESIMPULAN

Berdasarkan hasil dan pembahasan, maka diperoleh kesimpulan sebagai berikut:

1. Nilai penduga parameter dengan menggunakan metode ENPCR lebih baik dibandingkan nilai penduga parameter dengan menggunakan metode SPCR karena nilai penduga parameter ENPCR lebih mendekati nilai parameter sesungguhnya.
2. Nilai MMAD dan MSE metode ENPCR lebih kecil dibandingkan metode SPCR untuk seluruh kombinasi ukuran data n dan tingkat korelasi ρ .
3. Metode ENPCR lebih baik dibandingkan metode SPCR dalam menangani multikolinearitas berdasarkan kriteria MMAD dan MSE.

DAFTAR PUSTAKA

- Cook, R., & Forzani, L. 2008. Principal Fitted Components for Dimension Reduction in Regression. *Statistical Science*. **23**(4): 485-501.
- Draper, N. & H. Smith. 1992. *Analisis Regresi Terapan*. Ed. Ke-2. Terjemahan oleh Bambang Sumantri. Gramedia Pustaka Utama, Jakarta.
- Farhadi, Z., Belaghi, R., & Alma, O. 2019. Analysis of Penalized Regression Methods in a Simple Linear Model on the High-Dimensional Data. *American Journal of Theoretical and Applied Statistics*. **8**(5): 185-192.
- Friedman, J., Hestie, T. & Tibshirani, R. 2010. Reguralization Paths for generalized Linear Models Via Coordinate Descent. *Journal of Statistical Software*. **33**(1): 1-22.
- Gaspersz, V. 1991. *Ekonometri Terapan*. Ed. Ke-2. Tarsito, Bandung.
- Gujarati, D. N. & Porter, D. C. 2009. *Basic Econometrics*. McGraw-Hill Irwin, Boston.
- Hamada, A. A. 2024. Elastic Net Principal Component Regression With an Application. *International Journal of Science and Mathematics Education*. **1**(3): 13-23.
- Hucker, L., & Wahl, M. 2022. A Note on The Prediction Error of Principal Component Regression in High Dimensions. *Theory of Probability and Mathematical Statistics*. 1-25.
- Johnson, R. A. & Wichern, D. W. 2007. *Applied Multivariate Statistical Analysis*. Edisi 6. Pearson Prentice Hall Upper Saddle River, Amerika Serikat.

- Jolliffe, I. T. 2002. *Principal Component Analysis*. Second Edition. Springer Verlag, New York.
- Kutner, M. H., Nachtsheim, C. J., Neter, J. & Li, W. 2005. *Applied linear Statistic Model*. Ed. Ke-5. Mc-Graw hill, New York.
- Lee, H., Park, Y. M., & Lee, S. 2015. Principal Component Regression by Principal Component Selection. *Communications for Statistical Applications and Methods*. **22**(2): 173-180.
- Montgomery, D. G., Peck, E. A., & Vining, G. G. 2012. *Introduction to Linear Regression Analysis*. John Wiley & Sons Inc., New Jersey.
- Kawano, S., Fujisawa, H., Takada, T. & Shiroishi, T. 2015. Sparse Principal Component Regression with Adaptive Loading. *Computational Statistics & Data Analysis*. **89**: 192–203.
- Kim, J. 2019. Multicollinearity and Misleading Statistical Results. *Korean Journal of Anesthesiology*. **72**(6): 558 - 569.
- McClendon. & McKee J. 2002. *Multiple Regression and Causal Analysis*. Waveland Press, Amerika Serikat.
- Pratiwi, D. E., & Roosyanti, A. 2019. Analisis Faktor Penghambat Skripsi Mahasiswa Jurusan Pendidikan Guru Sekolah Dasar Universitas Wijaya Kusuma Surabaya. *Jurnal Pendidikan Dasar*. **10**(1): 101–114.
- Radiarta, I. N., Hasnawi, & Mustafa, A. 2013. Kondisi Kualitas Perairan Di Kabupaten Morowali Provinsi Sulawesi Tengah. *Jurnal Riset Akuakultur*. **8**(2): 299–309.
- Rougoor, C.W., Sundaram, R., & Van Arendonk, J. A. M. 2000. The Relation Between Breeding Management and 305-day Milk Production, Determined via Principal Components Regression and Partial Least Squares. *Livestock Product Science*. **66**(1):71-83.
- Sial, M. 2021. A Brief Introduction to Regression Analysis and Its Types. *Asian Journal of Probability and Statistics*. **13**(4): 58-63.

Sinaga, V., Safitri, D., & Rahmawati, R. 2019. Perbandingan Regresi Komponen Utama Dengan Regresi Kuadrat Terkecil Parsial Pada Indeks Pembangunan Manusia Provinsi Jawa Timur. *Jurnal Gaussian*. **8**(4): 496-505.

Sumodiningrat, G. 1998. *Ekonometrika Pengantar*. BPFE: Yogyakarta.

Zou, H., Hastie, T., & Tibshirani, R. 2006. Sparse Principal Component Analysis. *Journal of Computational and Graphical Statistics*. **15**(2): 265 - 286.