



**PENERAPAN *SPARSE PRINCIPAL COMPONENT ANALYSIS* (SPCA)
DAN *K-MEDOIDS* UNTUK PENGELOMPOKAN
KABUPATEN/KOTA BERDASARKAN INDIKATOR
KESEJAHTERAAN RAKYAT DI KALIMANTAN SELATAN**

SKRIPSI

**untuk memenuhi persyaratan
dalam menyelesaikan program sarjana Strata-1 Statistika**

**Oleh
M. FAYYAZ NAJMI
NIM. 2211017210023**

**PROGRAM STUDI S-1 STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMBUNG MANGKURAT
BANJARBARU
JULI 2026**



**PENERAPAN *SPARSE PRINCIPAL COMPONENT ANALYSIS* (SPCA)
DAN *K-MEDOIDS* UNTUK PENGELOMPOKAN
KABUPATEN/KOTA BERDASARKAN INDIKATOR
KESEJAHTERAAN RAKYAT DI KALIMANTAN SELATAN**

SKRIPSI

**untuk memenuhi persyaratan
dalam menyelesaikan program sarjana Strata-1 Statistika**

**Oleh
M. FAYYAZ NAJMI
NIM. 2211017210023**

**PROGRAM STUDI S-1 STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMBUNG MANGKURAT
BANJARBARU
JULI 2026**

SKRIPSI

PENERAPAN SPARSE PRINCIPAL COMPONENT ANALYSIS (SPCA) DAN K-MEDOIDS UNTUK PENGELOMPOKAN KABUPATEN/KOTA BERDASARKAN INDIKATOR KESEJAHTERAAN RAKYAT DI KALIMANTAN SELATAN

Oleh
M. FAYYAZ NAJMI
NIM. 2211017210023

Telah dipertahankan pada hari Senin, 6 Juli 2026 dan telah disetujui oleh dosen pembimbing dan dosen penguji sebagai berikut:

Pembimbing I



Selvi Annisa, S.Si., M.Si.
NIP. 199212262022032016

Penguji I



Yeni Rahkumawati, S.Mat., M.Si.
NIP. 199404032022032014

Pembimbing II



Rifqi Aulya Rahman, S.Mat., M.Si.
NIP. 199708072024061003

Penguji II



Al Hujjah Asianingrum, S.Stat M.Si.
NIP. 199601092024062003



Banjarbaru, Senin 13 Juli 2026
Mengetahui,
Koordinator Program Studi Statistika
FMIPA ULA

Dewi Anggraini, S.Si., M.App.Sci., Ph.D.
NIP. 198303282005012001

PERNYATAAN

Dengan ini saya menyatakan bahwa dalam skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka

Banjarbaru, 6 Juli 2026



M. Fayyaz Najmi
NIM. 2211017210023

ABSTRAK

Penerapan *Sparse Principal Component Analysis* (SPCA) dan *K-Medoids* untuk Pengelompokan Kabupaten/Kota Berdasarkan Indikator Kesejahteraan Rakyat di Kalimantan Selatan (Oleh: M. Fayyaz Najmi; Selvi Annisa dan Rifqi Aulya Rahman, 2026; 54 halaman)

Kesejahteraan rakyat merupakan salah satu tujuan utama pembangunan yang mencerminkan kualitas hidup masyarakat. Pengukuran kesejahteraan rakyat tidak dapat dilakukan hanya berdasarkan satu bidang, melainkan memerlukan berbagai bidang yang mencakup aspek pendidikan, kesehatan, fertilitas dan keluarga berencana, perumahan, pemanfaatan teknologi informasi, serta program bantuan pemerintah. Banyaknya bidang yang digunakan menyebabkan data bersifat multidimensional dan saling berkaitan sehingga diperlukan metode yang mampu menyederhanakan struktur data tanpa menghilangkan informasi penting. Penelitian ini bertujuan menerapkan *Sparse Principal Component Analysis* (SPCA) untuk mereduksi dimensi indikator kesejahteraan rakyat serta mengelompokkan kabupaten/kota di Provinsi Kalimantan Selatan menggunakan metode *K-Medoids* berdasarkan skor komponen utama yang dihasilkan. Data yang digunakan merupakan data indikator kesejahteraan rakyat kabupaten/kota di Provinsi Kalimantan Selatan tahun 2024 yang bersumber dari Survei Sosial Ekonomi Nasional (SUSENAS) yang dipublikasikan oleh Badan Pusat Statistik. SPCA diterapkan menggunakan pendekatan *Elastic Net* untuk menghasilkan komponen utama yang lebih sederhana. Hasil SPCA menunjukkan bahwa dari 89 variabel awal diperoleh 8 variabel terpilih yang mampu menjelaskan 84.36% keragaman data. Skor komponen utama yang dihasilkan kemudian digunakan sebagai input pada metode *K-Medoids*. Hasil pengelompokan menghasilkan tiga kluster optimal dengan nilai *Silhouette Coefficient* sebesar 0.136. Kluster 1 terdiri atas 5 kabupaten/kota dengan tingkat kesejahteraan sedang, kluster 2 terdiri atas 7 kabupaten/kota dengan tingkat kesejahteraan rendah, dan kluster 3 terdiri atas 1 kabupaten/kota dengan tingkat kesejahteraan tinggi. Dengan demikian, kombinasi SPCA dan *K-Medoids* mampu mereduksi kompleksitas indikator kesejahteraan rakyat serta mengidentifikasi kelompok kabupaten/kota berdasarkan karakteristik kesejahteraan rakyat di Provinsi Kalimantan Selatan.

Kata kunci: Kesejahteraan Rakyat; *Sparse Principal Component Analysis*; *K-Medoids*; Pengelompokan; Kalimantan Selatan.

ABSTRACT

Application of Sparse Principal Component Analysis (SPCA) and K-Medoids for Clustering Regencies/Municipalities Based on People's Welfare Indicators in South Kalimantan Province (By: M. Fayyaz Najmi; Selvi Annisa and Rifqi Aulya Rahman, 2026; 54 pages)

People's welfare is one of the main goals of development that reflects the quality of life of the community. Measuring people's welfare cannot be based on a single indicator alone, but requires various indicators covering aspects of education, health, fertility and family planning, housing, utilization of information technology, and government assistance programs. The large number of indicators used causes the data to be multidimensional and interrelated, necessitating a method capable of simplifying the data structure without losing important information. This study aims to apply Sparse Principal Component Analysis (SPCA) to reduce the dimensionality of people's welfare indicators and to cluster regencies/municipalities in South Kalimantan Province using the K-Medoids method based on the resulting principal component scores. The data used are people's welfare indicator data for regencies/municipalities in South Kalimantan Province in 2024, sourced from the National Socioeconomic Survey (SUSENAS) published by the Central Bureau of Statistics. SPCA was applied using the Elastic Net approach to produce simpler principal components. The SPCA results indicate that the initial 89 variables were represented by eight sparse principal components, which collectively explained 84.36% of the total data variance. The resulting principal component scores were then used as input for the K-Medoids method. The clustering results yielded three optimal clusters with a Silhouette Coefficient value of 0.136. Cluster 1 consists of 5 regencies/municipalities with a moderate welfare level, Cluster 2 consists of 7 regencies/municipalities with a low welfare level, and Cluster 3 consists of 1 regencies/municipalities with a high welfare level. Thus, the combination of SPCA and K-Medoids is capable of reducing the complexity of people's welfare indicators and identifying groups of regencies/municipalities based on welfare characteristics in South Kalimantan Province.

Keywords: People's Welfare; Sparse Principal Component Analysis; K-Medoids; Clustering; South Kalimantan.

PRAKATA

Puji syukur kehadiran Allah SWT atas rahmat, taufik, hidayah, dan inayah-Nya sehingga penulis dapat menyelesaikan penulisan Tugas Akhir dengan judul “Penerapan *Sparse Principal Component Analysis* (SPCA) dan *K-Medoids* untuk Pengelompokan kabupaten/kota Berdasarkan Indikator Kesejahteraan Rakyat di Kalimantan Selatan”. Penulisan ini dilakukan dengan tujuan memenuhi persyaratan dalam menyelesaikan program sarjana di Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lambung Mangkurat. Penulis menyadari adanya kesulitan dan hambatan dari segi penulisan maupun penyusunan kata-kata yang belum tersusun dengan baik dan hal lainnya. Namun, berkat dukungan dan bimbingan berbagai pihak, penulisan ini dapat diselesaikan dengan baik. Dengan tulus, penulis mengucapkan terima kasih kepada pihak-pihak yang telah membantu, khususnya kepada:

1. Orang tua dan keluarga yang selalu memberikan dukungan serta mendoakan untuk penyelesaian Tugas Akhir;
2. Ibu Selvi Annisa, S.Si., M.Si. dan Bapak Rifqi Aulya Rahman, S.Mat., M.Si. selaku dosen pembimbing yang telah bersedia meluangkan waktu, tenaga, dan pikiran dalam pelaksanaan penelitian serta penyelesaian Tugas Akhir;
3. Ibu Yeni Rahkmawati, S.Mat., M.Si. dan Ibu Al Hujjah Asianingrum, S.Stat M.Si. selaku dosen penguji yang telah memberikan masukan, saran, dan bantuan dalam perbaikan penulisan Tugas Akhir;
4. Koordinator Program Studi Statistika FMIPA ULM beserta seluruh dosen dan staf yang selalu memberikan ilmu, motivasi, nasihat, dan mendukung selama masa perkuliahan;
5. Teman-teman yang senantiasa memberikan dukungan dan semangat selama proses penelitian dan penyusunan Tugas Akhir, khususnya “keluarga Kopi Kenangan”, “Funtastic”, dan anggota grup “ANOVA”;

Penulis menyadari bahwa dalam penyusunan Tugas Akhir ini masih terdapat berbagai kekurangan. Oleh karena itu, penulis sangat mengharapkan kritik dan saran yang membangun dari semua pihak demi perbaikan dan penyempurnaan Tugas Akhir ini. Semoga segala bantuan, dukungan, dan kebaikan yang telah diberikan oleh semua pihak mendapatkan balasan terbaik dari Tuhan Yang Maha Esa. Penulis berharap semoga Tugas Akhir ini dapat memberikan manfaat, menambah wawasan, serta menjadi referensi yang berguna bagi pembaca dan penelitian selanjutnya.

Banjarbaru, 6 Juli 2026

M. Fayyaz Najmi

DAFTAR ISI

HALAMAN JUDUL	i
LEMBAR PENGESAHAN	ii
PERNYATAAN	iii
ABSTRAK.....	iv
ABSTRACT	v
PRAKATA.....	vi
DAFTAR ISI	vii
DAFTAR GAMBAR.....	ix
DAFTAR TABEL.....	x
DAFTAR LAMPIRAN.....	xi
DAFTAR LAMBANG, DAN SINGKATAN.....	xii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah	3
1.3 Tujuan Penelitian.....	4
1.4 Manfaat Penelitian	4
BAB II TINJAUAN PUSTAKA	5
2.1 Kajian Penelitian Terdahulu	5
2.2 Kajian Teori.....	7
2.2.1 <i>Exploratory Data Analysis</i> (EDA)	7
2.2.2 Multikolinearitas	8
2.2.3 Standarisasi Data.....	9
2.2.4 <i>Elastic Net</i>	10
2.2.5 <i>Sparse Principal Component Analysis</i> (SPCA)	11
2.2.6 <i>K-Medoids</i>	18
2.2.7 <i>Silhouette Coefficient</i>	20
2.2.8 <i>Relative Standard Error</i> (RSE).....	20
2.2.9 <i>Kesejahteraan Rakyat</i>	21
BAB III METODE PENELITIAN.....	28
3.1 Sumber Data	28
3.2 Prosedur Penelitian	28
BAB IV HASIL DAN PEMBAHASAN	33
4.1 Persiapan data	33
4.2 <i>Exploratory Data Analysis</i> (EDA)	33
4.3 Pemeriksaan Multikolinearitas	35
4.4 Standarisasi Data	36
4.5 Proses <i>Sparse Principal Component Analysis</i>	37
4.6 Analisis Kluster	41
4.6.1 Penentuan Jumlah Kluster Optimal.....	41
4.6.2 Pengelompokan Menggunakan Metode <i>K-Medoids</i>	42
4.6.3 Kluster yang Terbentuk	45
4.6.4 Pemilihan Variabel Representatif.....	46
4.6.5 Karakterisasi Kluster	47
BAB V PENUTUP	56

5.1. Kesimpulan	56
5.2. Saran.....	57
DAFTAR PUSTAKA	58
LAMPIRAN.....	61
RIWAYAT HIDUP.....	97

PRODI STATISTIKA

DAFTAR GAMBAR

Gambar 2.1 Ilustrasi <i>Box Plot</i>	8
Gambar 2.2 Ilustrasi Proses <i>K-Medoids</i>	19
Gambar 3.1 Diagram Alir	32
Gambar 4.1 Distribusi 5 Variabel dengan Rentang Nilai Terbesar	34
Gambar 4.2 Nilai Silhouette Coefficient	42
Gambar 4.3 Peta Tematik Hasil Klasterisasi.....	45

PRODI STATISTIKA

DAFTAR TABEL

Tabel 2.1 Kajian Penelitian Terdahulu	5
Tabel 3.1 Bidang dan Kategori Indikator Kesejahteraan Rakyat yang Digunakan dalam Penelitian.....	28
Tabel 4.1 Statistika Deskriptif 5 Variabel Dengan Nilai <i>Range</i> Terbesar	34
Tabel 4.2 Perhitungan Korelasi	35
Tabel 4.3 Data untuk Kota Banjarmasin	36
Tabel 4.4 Data Kota Banjarmasin Setelah di Standarisasi.....	36
Tabel 4.5 Hasil SVD PCA	37
Tabel 4.6 Matriks Bobot Awal A Sebelum Penalti	38
Tabel 4.7 Hasil proses iterasi B	39
Tabel 4.8 Nilai Variansi dan PEV SKU.....	40
Tabel 4.9 Skor Sparse Komponen Utama.....	41
Tabel 4.10 Nilai Medoid Awal.....	42
Tabel 4.11 Hasil Pengelompokan Iterasi Pertama	43
Tabel 4.12 Hasil Pengelompokan Berdasarkan Jarak Minimum	44
Tabel 4.13 Hasil Klaster Akhir	45
Tabel 4.14 Variabel Representatif Berdasarkan Bobot Tertinggi.....	46
Tabel 4.15 Karakteristik Pendidikan.....	47
Tabel 4.16 Karakteristik Kesehatan.....	48
Tabel 4.17 Karakteristik Fertilitas dan KB.....	49
Tabel 4.18 Karakteristik Perumahan	50
Tabel 4.19 Karakteristik Pemanfaatan Teknologi Informasi	50
Tabel 4.20 Karakteristik Bantuan/Program Pemerintah.....	51
Tabel 4.21 Karakteristik Variabel Pembentuk Klaster Kesejahteraan Rakyat	52

DAFTAR LAMPIRAN

Lampiran 1. Definisi Operasional Data Penelitian	61
Lampiran 2. Data yang digunakan.....	72
Lampiran 3. Variabel yang Berkolerasi Tinggi	73
Lampiran 4. Data Yang Distandarisasi.....	75
Lampiran 5. Tuning Parameter LASSO	76
Lampiran 6. <i>Syntax Program R</i>	78

PRODI STATISTIKA

DAFTAR LAMBANG, DAN SINGKATAN

x_{ij}	=	Data ke- i pada variabel ke- j
$\bar{x}_j$	=	Nilai rata-rata pada variabel ke- j
s_j	=	Nilai simpangan baku pada variabel ke- j
z_{ij}	=	Data ke- i hasil standarisasi pada variabel ke- j
A	=	Matriks <i>loading</i> PCA
X	=	Matriks data
KU	=	Komponen utama
SKU	=	<i>Sparse</i> Komponen utama
$\beta^{(f)}$	=	Nilai <i>loading</i> pada iterasi ke- f
y	=	Variabel respon pada formulasi <i>Elastic Net</i>
ℓ_1	=	Parameter penalti LASSO
ℓ_2	=	Parameter penalti <i>Ridge</i>
Q	=	Matriks orthogonal hasil dekomposisi QR
R	=	Matriks segitiga atas hasil dekomposisi QR
σ_k	=	Nilai singular ke- k
C_i	=	Klaster ke- i
m_i	=	<i>Medoid</i> klaster ke- i
d	=	Jarak <i>Euclidean</i> antara objek dan <i>medoid</i>
$a(i)$	=	Rata-rata jarak data terhadap klaster yang sama
$b(i)$	=	Rata-rata jarak data terhadap klaster terdekat
$s(i)$	=	Nilai <i>Silhouette Coefficient</i>
n	=	Jumlah observasi
p	=	Jumlah variabel
LASSO	=	<i>Least Absolute Shrinkage and Selection Operator</i>
PCA	=	<i>Principal Component Analysis</i>
SPCA	=	<i>Sparse Principal Component Analysis</i>