

***PRINCIPAL COMPONENT ANALYSIS (PCA) BILOT DAN K-MEANS PADA
PEMETAAN WILAYAH DI PROVINSI LAMPUNG BERDASARKAN
INDIKATOR KESEJAHTERAAN MASYARAKAT***

(Skripsi)

Oleh

**TIKA AYUWARDANI
1917031081**



**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2026**

ABSTRACT

PRINCIPAL COMPONENT ANALYSIS (PCA) BILOT DAN K-MEANS PADA PEMETAAN WILAYAH DI PROVINSI LAMPUNG BERDASARKAN INDIKATOR KESEJAHTERAAN MASYARAKAT

By

TIKA AYUWARDANI

Community welfare is one of the primary indicators used to measure the success of regional development. Differences in health, education, economic, and social conditions contribute to welfare disparities among regencies and cities in Lampung Province. This study applies *Principal Component Analysis* (PCA) Biplot and *K-Means* clustering to map regions based on 10 community welfare indicators across 15 regencies/cities in Lampung Province in 2025. The results show that the two principal components formed are able to explain 71.22% of the total data variability. The first principal component represents health, education, and economic aspects, while the second principal component represents employment, adequate housing, and crime-related aspects. The *K-Means* analysis produced two optimal clusters: Cluster 1 consists of 13 regencies/cities with relatively lower welfare levels, while Cluster 2 consists of Bandar Lampung City and Metro City, which have relatively higher welfare levels. The findings indicate clear differences in welfare characteristics among regions and can serve as a basis for formulating more effective and targeted development policies.

Keywords: Community Welfare, PCA Biplot, *K-Means*, Regional Mapping, Lampung Province

ABSTRAK

PRINCIPAL COMPONENT ANALYSIS (PCA) BILOT DAN K-MEANS PADA PEMETAAN WILAYAH DI PROVINSI LAMPUNG BERDASARKAN INDIKATOR KESEJAHTERAAN MASYARAKAT

Oleh

TIKA AYUWARDANI

Kesejahteraan masyarakat merupakan salah satu indikator utama dalam mengukur keberhasilan pembangunan daerah. Perbedaan kondisi kesehatan, pendidikan, ekonomi, dan sosial menyebabkan terjadinya ketimpangan kesejahteraan antar kabupaten/kota di Provinsi Lampung. Penelitian ini menerapkan metode *Principal Component Analysis (PCA) Bilot* dan *K-Means* untuk memetakan wilayah berdasarkan 10 indikator kesejahteraan masyarakat pada 15 kabupaten/kota di Provinsi Lampung tahun 2025. Hasil analisis menunjukkan bahwa dua komponen utama yang terbentuk mampu menjelaskan 71,22% keragaman data. Komponen utama pertama merepresentasikan aspek kesehatan, pendidikan, dan ekonomi, sedangkan komponen utama kedua merepresentasikan aspek ketenagakerjaan, hunian layak, dan kriminalitas. Analisis *K-Means* menghasilkan dua klaster optimal, yaitu Klaster 1 yang terdiri atas 13 kabupaten/kota dengan tingkat kesejahteraan relatif lebih rendah dan Klaster 2 yang terdiri atas Kota Bandar Lampung dan Kota Metro dengan tingkat kesejahteraan relatif lebih tinggi. Hasil penelitian menunjukkan adanya perbedaan karakteristik kesejahteraan antarwilayah yang cukup jelas, sehingga dapat menjadi dasar dalam perumusan kebijakan pembangunan yang lebih efektif dan tepat sasaran.

Kata Kunci: Kesejahteraan Masyarakat, PCA Bilot, *K-Means*, Pemetaan Wilayah, Provinsi Lampung.

***PRINCIPAL COMPONENT ANALYSIS (PCA) BILOT DAN K-MEANS PADA
PEMETAAN WILAYAH DI PROVINSI LAMPUNG BERDASARKAN
INDIKATOR KESEJAHTERAAN MASYARAKAT***

Oleh

TIKA AYUWARDANI

Skripsi

Sebagai Salah Satu Syarat untuk Mencapai Gelar
SARJANA MATEMATIKA

pada

Jurusan Matematika
Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Lampung



FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2026

Judul Skripsi : **PRINCIPAL COMPONENT ANALYSIS (PCA) BILOT DAN K-MEANS PADA PEMETAAN WILAYAH DI PROVINSI LAMPUNG BERDASARKAN INDIKATOR KESEJAHTERAAN MASYARAKAT**

Nama Mahasiswa : **Tika Ayuwardani**

Nomor Pokok Mahasiswa : 1917031081

Jurusan : Matematika

Fakultas : Matematika dan Ilmu Pengetahuan Alam



1. Komisi Pembimbing

Dr. Khoirin Nisa, S.Si., M.Si.
NIP.197407262000032001

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

2. Ketua Jurusan Matematika

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

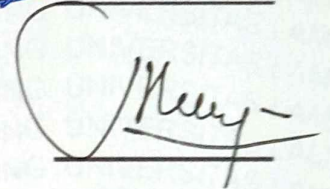
MENGESAHKAN

1. Tim Penguji

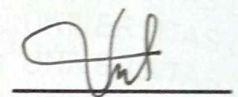
Ketua : **Dr. Khoirin Nisa, S.Si., M.Si.**



Sekretaris : **Dr. Aang Nuryaman, S.Si., M.Si.**



Penguji
Bukan Pembimbing : **Drs. Nusyirwan, M.Si.**



2. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam



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

NIP. 197110012005011002

Tanggal Lulus Ujian Skripsi : **15 Juni 2026**

SURAT PERNYATAAN SKRIPSI MAHASISWA

Yang bertanda tangan dibawah ini :

Nama : Tika Ayuwardani
Nomor Pokok Mahasiswa : 1917031081
Jurusan : Matematika
Judul Skripsi : ***PRINCIPAL COMPONENT ANALYSIS (PCA) BILOT DAN K-MEANS PADA PEMETAAN WILAYAH DI PROVINSI LAMPUNG BERDASARKAN INDIKATOR KESEJAHTERAAN MASYARAKAT***

Dengan ini menyatakan bahwa penelitian ini adalah hasil pekerjaan saya sendiri dan apabila dikemudian hari hasil penelitian atau tugas akhir saya merupakan salinan atau dibuat oleh orang lain, maka saya bersedia menerima sanksi sesuai dengan ketentuan akademik yang berlaku. Semua tulisan yang tertuang dalam skripsi ini telah mengikuti kaidah penulisan karya tulis ilmiah Universitas Lampung.

Bandar Lampung, 15 Juni 2026
Yang membuat pernyataan,



Tika Ayuwardani

RIWAYAT HIDUP

Penulis bernama lengkap Tika Ayuwardani, lahir di Lampung Utara pada tanggal 21 Oktober 2000, sebagai putri dari pasangan Bapak Supriyanto dan Ibu Jumiyati. Penulis merupakan anak bungsu dari tiga bersaudara.

Penulis menempuh pendidikan di Sekolah Dasar Negeri (SDN) 1 Jati Indah pada tahun 2007-2013, Sekolah Menengah Pertama Negeri (SMPN) 1 Tanjung Bintang pada tahun 2013-2016, dan Sekolah Menengah Atas Negeri (SMAN) 1 Tanjung Bintang pada tahun 2016-2019. Pada tahun 2019, penulis terdaftar menjadi mahasiswi Program Studi S1 Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam (FMIPA) Universitas Lampung melalui jalur SBMPTN.

Selama menempuh pendidikan di perguruan tinggi, penulis pernah mengikuti kegiatan organisasi kemahasiswaan, di antaranya sebagai anggota Koperasi Mahasiswa Universitas Lampung periode 2021/2022 dan anggota UKM-F ROIS FMIPA Universitas Lampung periode 2020. Sebagai bentuk penerapan ilmu yang diperoleh selama perkuliahan, penulis melaksanakan Kerja Praktik di Dinas Kehutanan Provinsi Lampung pada tanggal 4 Januari sampai 12 Februari 2022. Selanjutnya, pada tanggal 29 Juni sampai 7 Agustus 2022, penulis melaksanakan Kuliah Kerja Nyata (KKN) di Desa Sukanegeri Jaya, Kecamatan Talang Padang, Kabupaten Tanggamus, sebagai wujud pengabdian kepada masyarakat dalam rangka pelaksanaan Tri Dharma Perguruan Tinggi.

KATA INSPIRASI

“Sesungguhnya bersama kesulitan ada kemudahan.”

(Q.S. Al-Insyirah: 5-6)

"Allah tidak membebani seseorang melainkan sesuai dengan kesanggupannya."

(QS. Al-Baqarah: 286)

“Ya Rabbku, lapangkanlah untukku dadaku, dan mudahkanlah untukku urusanku, dan lepaskanlah kekakuan dari lidahku, supaya mereka mengerti perkataanku.”

(Q.S Thoha: 25-28)

"It does not matter how slowly you go as long as you do not stop."

(Confucius)

“It always seems impossible until it's done.”

(Nelson Mandela)

PERSEMBAHAN

Alhamdulillah, Puji dan syukur kepada Allah SWT yang selalu memberi penguatan dalam setiap langkah yang ku tempuh sehingga segala bentuk proses yang ditemui dapat dilalui dengan penuh makna.

Oleh karena itu, dengan rasa syukur dan bahagia saya persembahkan rasa terima kasih saya kepada:

Orang Tua Tercinta

Terima kasih atas segala doa yang tidak pernah putus, kasih sayang yang tulus, serta dukungan yang senantiasa diberikan dalam setiap langkah kehidupan penulis. Terima kasih atas segala pengorbanan, nasihat, dan kepercayaan yang telah menjadi kekuatan bagi penulis untuk menyelesaikan pendidikan ini.

Dosen Pembimbing dan Dosen Pembahas

Terima kasih kepada dosen pembimbing dan pembahas yang sudah sangat membantu, memberikan motivasi, memberikan arahan serta ilmu yang berharga.

Teman dan Sahabat

Terima kasih kepada semua orang-orang baik yang telah memberikan pengalaman, semangat, motivasi, dan selalu ada saat suka maupun duka.

SANWACANA

Puji dan syukur penulis ucapkan kehadiran Allah SWT yang telah memberikan rahmat dan hidayah-Nya, sehingga penulis dapat menyelesaikan skripsi yang berjudul "*Principal Component Analysis (PCA) Biplot dan K-Means pada Pemetaan Wilayah di Provinsi Lampung Berdasarkan Indikator Kesejahteraan Masyarakat*" sebagai salah satu persyaratan untuk meraih gelar strata satu (S1) di Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.

Penyelesaian karya ilmiah ini tidak terlepas dari bantuan, bimbingan, serta doa dari berbagai pihak. Oleh karena itu, pada kesempatan ini penulis menyampaikan ucapan terima kasih yang sebesar-besarnya kepada:

1. Bapak Dr. Eng. Heri Satria, S.Si., M.Si., selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam (FMIPA) Universitas Lampung.
2. Ibu Dr. Khoirin Nisa, S.Si., M.Si., selaku Dosen Pembimbing I sekaligus Dosen Pembimbing Akademik, yang senantiasa memberikan bimbingan, arahan, saran serta bantuan kepada penulis dalam menyelesaikan skripsi ini.
3. Bapak Dr. Aang Nuryaman, S.Si., M.Si., selaku Dosen Pembimbing II sekaligus Ketua Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam (FMIPA) Universitas Lampung, yang telah memberikan bimbingan, saran, dan masukan kepada penulis dalam penyelesaian skripsi ini.
4. Bapak Drs. Nusyirwan, M.Si., selaku dosen pembahas, yang telah meluangkan waktu untuk membahas, memberikan kritik, saran, serta evaluasi yang berharga kepada penulis dalam penyelesaian skripsi ini.
5. Seluruh dosen, staff, dan karyawan Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam (FMIPA) Universitas Lampung.

6. Orang tua tercinta, Mamak, Pak Wi, dan Bapak, yang senantiasa memberikan doa, dukungan, kasih sayang, serta pengorbanan yang tiada henti demi keberhasilan penulis. Penulis berharap dapat membahagiakan serta menjadi kebanggaan kalian di kemudian hari.
7. Kakak-kakak tersayang, khususnya Mba Ten, yang telah memberikan dukungan, motivasi, serta doa kepada penulis.
8. Berlian, Mba Linda, Mega, Zida, dan Sinta, yang telah memberikan bantuan, dukungan, serta kebersamaan yang berharga selama masa perkuliahan.
9. Teman-teman seperjuangan, Siti, Yuni, Lisna, dan Karima, yang selalu memberikan bantuan, semangat, dan kebersamaan selama proses penyusunan skripsi ini.
10. Teman-teman Matematika 2019, terimakasih atas kebersamaannya.
11. Almamater tercinta, Universitas Lampung.
12. Seluruh pihak yang telah membantu penulis dalam menyelesaikan skripsi ini yang tidak dapat disebutkan satu persatu.

Bandar Lampung, 15 Juni 2026

Penulis

Tika Ayuwardani

DAFTAR ISI

	Halaman
DAFTAR TABEL	i
DAFTAR GAMBAR.....	ii
I. PENDAHULUAN.....	1
1.1 Latar Belakang dan Masalah	1
1.2 Tujuan Penelitian	3
1.3 Manfaat Penelitian	3
II. TINJAUAN PUSTAKA.....	4
2.1 Analisis Multivariat	4
2.2 <i>Principal Component Analysis</i> (PCA)	5
2.2.1 Menentukan Banyaknya Komponen Utama	5
2.2.2 Koefisien Komponen Utama.....	6
2.3 Analisis Biplot.....	7
2.3.1 Definisi Analisis Biplot.....	7
2.3.2 Penguraian Nilai Singular (<i>Singular Value Decomposition</i>)	8
2.3.3 Interpretasi Biplot.....	10
2.3.4 Ukuran Kesesuaian Biplot	11
2.4 <i>K-Means Cluster</i>	12
2.5 Standarisasi Data.....	13
2.6 Kesejahteraan Masyarakat	14
III. METODOLOGI PENELITIAN	17
3.1 Waktu dan Tempat Penelitian	17
3.2 Data Penelitian	17
3.3 Metode Penelitian	18
IV. HASIL DAN PEMBAHASAN	19
4.1 Analisis Deskriptif	19
4.2 Standarisasi Data.....	20
4.3 <i>Principal Component Analysis</i> (PCA)	21

4.3.1	Menentukan Komponen Utama	21
4.3.2	Nilai Koefisien Komponen Utama (<i>Loading</i>).....	22
4.3.3	Transformasi Skor Komponen Utama	24
4.4	Analisis Biplot.....	25
4.4.1	Interpretasi Biplot.....	27
4.4.2	Ukuran Kesesuaian Biplot	29
4.5	Analisis <i>K-Means</i>	29
4.5.1	Menentukan Jumlah Kluster Optimum	29
4.5.2	Pembentukan Kluster Optimum	30
4.5.3	Interpretasi Karakterisasi Kluster	32
V.	KESIMPULAN	34
	DAFTAR PUSTAKA	35
	LAMPIRAN.....	37

DAFTAR TABEL

Tabel	Halaman
1. Variabel Indikator Kesejahteraan Masyarakat	17
2. Analisis Deskriptif Data.....	19
3. Hasil Standarisasi Data	20
4. Nilai Eigen dan Proporsi Kumulatif.....	21
5. Nilai Koefisien Komponen Utama.....	23
6. Hasil Pengelompokan Variabel pada Komponen Utama.....	23
7. Transformasi Skor Komponen Utama	25
8. Panjang Vektor PCA Biplot.....	27
9. Hasil Pembentukan Kluster Optimum.....	31
10. Karakterisasi Kluster	32

DAFTAR GAMBAR

Gambar	Halaman
1. <i>Scree Plot</i> dari Nilai Eigen.....	22
2. PCA Biplot.....	26
3. Plot <i>Silhouette Coefficient</i>	30
4. Plot Hasil Pembentukan Klaster	31

I. PENDAHULUAN

1.1 Latar Belakang dan Masalah

Kesejahteraan masyarakat merupakan salah satu indikator penting dalam mengukur keberhasilan pembangunan suatu negara. Hal ini didasarkan pada pengertian bahwa kesejahteraan masyarakat adalah kondisi terpenuhinya kebutuhan dasar masyarakat secara layak, baik dari aspek ekonomi, kesehatan, pendidikan, maupun kualitas hidup secara menyeluruh (Suparlan, 2004). Di Indonesia, pemerintah terus berupaya meningkatkan kesejahteraan melalui berbagai program di bidang pendidikan, kesehatan, ekonomi, dan infrastruktur.

Meskipun menunjukkan berbagai kemajuan, ketimpangan kesejahteraan antarwilayah dan antarkelompok sosial masih menjadi tantangan besar. Menurut Badan Pusat Statistik Indonesia (2023), angka kemiskinan nasional mengalami penurunan dari tahun ke tahun. Namun demikian, kesenjangan antarprovinsi dalam akses terhadap pendidikan, layanan kesehatan, dan pendapatan masyarakat masih tergolong tinggi. Hal ini mencerminkan adanya ketidakseimbangan pencapaian kesejahteraan di berbagai daerah di Indonesia.

Provinsi Lampung, sebagai salah satu provinsi di Pulau Sumatera yang memiliki potensi sumber daya alam dan manusia yang besar, juga menghadapi persoalan ketimpangan kesejahteraan antar kabupaten/kota. Menurut Badan Pusat Statistik Lampung (2023), beberapa daerah di Lampung masih menghadapi tantangan dalam pemenuhan kebutuhan dasar seperti layanan kesehatan, pemerataan

pendidikan, dan tingkat kemiskinan yang relatif tinggi, terutama di wilayah pedesaan.

Penelitian sebelumnya yang dilakukan oleh Widyatama (2017) menunjukkan adanya kesenjangan kesejahteraan antarwilayah di Provinsi Lampung, yang ditunjukkan melalui indikator seperti Indeks Pembangunan Manusia (IPM), pendapatan per kapita, dan persentase penduduk miskin. Hal ini menunjukkan pentingnya dilakukan kajian lebih lanjut guna memahami pola persebaran kesejahteraan masyarakat secara menyeluruh.

Salah satu pendekatan statistik yang dapat digunakan untuk menggambarkan dan menganalisis pola tersebut adalah *Principal Component Analysis* (PCA) Biplot dan *K-Means*. Metode ini efektif dalam mereduksi dimensi data multivariat tanpa kehilangan banyak informasi penting, serta mampu memberikan visualisasi secara simultan antara objek (wilayah) dan variabel (indikator kesejahteraan). Metode ini telah diterapkan dalam berbagai penelitian sebelumnya, seperti oleh Hutagalung, R. & Nuramaliyah (2026) yang menggunakan *Principal Component Analysis* (PCA) dan *K-Means* untuk pengelompokan kabupaten/kota di Sumatera Utara berdasarkan indikator infrastruktur dan layanan dasar, serta oleh Rosyada, I. A. & Utari, D. T. (2024) yang menerapkan *Principal Component Analysis* (PCA) untuk reduksi variabel pada Algoritma *K-Means Clustering*.

Berdasarkan pemaparan di atas, peneliti bermaksud untuk mengetahui bagaimana pola persebaran wilayah di Provinsi Lampung berdasarkan indikator kesejahteraan masyarakat, serta mengidentifikasi faktor-faktor dominan yang memengaruhinya. Oleh karena itu, peneliti tertarik untuk melakukan pengelompokan data menggunakan *Principal Component Analysis* (PCA) Biplot dan *K-Means* dalam penelitian yang berjudul “*Principal Component Analysis* (PCA) Biplot dan *K-Means* pada Pemetaan Wilayah di Provinsi Lampung Berdasarkan Indikator Kesejahteraan Masyarakat”.

1.2 Tujuan Penelitian

Adapun tujuan dalam penelitian ini adalah untuk mengidentifikasi pola persebaran wilayah di Provinsi Lampung berdasarkan indikator kesejahteraan masyarakat, serta menentukan faktor-faktor dominan yang memengaruhi tingkat kesejahteraan antarwilayah dengan menggunakan *Principal Component Analysis* (PCA) Biplot dan *K-Means*.

1.3 Manfaat Penelitian

Adapun manfaat dalam penelitian ini, yaitu menambah pengetahuan dan wawasan mengenai penerapan analisis biplot komponen utama untuk pemetaan wilayah berdasarkan indikator kesejahteraan masyarakat. Selain itu, hasil penelitian ini juga diharapkan dapat menjadi referensi dan bahan pertimbangan dalam merumuskan kebijakan yang lebih tepat sasaran untuk mengatasi ketimpangan kesejahteraan antarwilayah di Provinsi Lampung.

II. TINJAUAN PUSTAKA

2.1 Analisis Multivariat

Menurut Hair, *et al.* (2010), analisis multivariat adalah himpunan metode statistik yang digunakan untuk mengobservasi dan menganalisis lebih dari satu variabel pengukuran pada waktu yang bersamaan. Tujuan utamanya adalah untuk memahami struktur data dan hubungan antar variabel secara lebih menyeluruh. Menurut Johnson & Wichern (2007), data multivariat merupakan jenis data yang terdiri dari lebih dari satu variabel yang diamati atau diukur secara simultan pada masing-masing objek atau unit pengamatan. Setiap objek dalam data multivariat memiliki nilai pada beberapa variabel sekaligus, sehingga memberikan informasi yang lebih kaya dibandingkan data univariat (satu variabel) atau bivariat (dua variabel).

Menurut Rencher (2002), data multivariat dapat direpresentasikan dalam bentuk matriks data berdimensi $n \times p$, di mana n adalah jumlah objek pengamatan dan p adalah jumlah variabel. Setiap baris dalam matriks tersebut mewakili satu objek pengamatan, sedangkan setiap kolom mewakili satu variabel. Pendekatan ini memungkinkan analisis hubungan, pola, dan struktur dalam data yang kompleks.

2.2 *Principal Component Analysis (PCA)*

Principal Component Analysis (PCA) atau Analisis Komponen Utama (AKU) pertama kali diperkenalkan oleh Karl Pearson pada tahun 1901 dan dikembangkan lebih lanjut oleh Harold Hotelling pada tahun 1933. Menurut Jolliffe (2002), Analisis Komponen Utama merupakan metode yang digunakan untuk mereduksi dimensi data dengan cara mengubah variabel-variabel asli yang saling berkorelasi menjadi sejumlah variabel baru yang saling tidak berkorelasi (*orthogonal*) satu sama lain, yang disebut sebagai komponen utama (*principal components*).

Menurut Johnson & Wichern (2007), komponen utama dalam analisis PCA dibentuk sebagai kombinasi linear dari variabel-variabel asli, dengan tujuan untuk merangkum sebanyak mungkin varians dalam data. Pembentukan komponen utama mengikuti sejumlah ketentuan dasar yang penting, yaitu:

1. Terdapat sebanyak p komponen utama yang merupakan kombinasi linear dari p variabel asli.
2. Setiap komponen utama saling *ortogonal* (tegak lurus) satu sama lain dan saling bebas (*independen*).
3. Komponen utama tidak berkorelasi satu sama lain dan disusun berdasarkan urutan varians dari yang terbesar ke yang terkecil.

Karakteristik ini menjadikan PCA sebagai metode yang efektif untuk reduksi dimensi data tanpa kehilangan banyak informasi penting, serta sangat berguna dalam eksplorasi data multivariat dan visualisasi struktur data

2.2.1 **Menentukan Banyaknya Komponen Utama**

Penentuan jumlah komponen utama bertujuan untuk memperoleh representasi data yang ringkas, dengan tetap mempertahankan sebanyak mungkin informasi yang terkandung dalam variabel-variabel asli. Menurut Jolliffe (2002), jumlah komponen utama yang dipilih harus cukup untuk menjelaskan variansi data secara

signifikan, namun tidak berlebihan sehingga menghasilkan kompleksitas yang tidak perlu. Berikut adalah metode-metode yang umum digunakan untuk menentukan banyaknya komponen utama:

1. Kriteria Kaiser (*Eigenvalue* ≥ 1)

Menurut Hair, *et al.* (2010), metode ini menyatakan bahwa hanya komponen utama dengan nilai eigen (*eigenvalue*) ≥ 1 yang dianggap signifikan dan layak dipertahankan, karena komponen tersebut mampu menjelaskan varians yang lebih besar dibandingkan satu variabel asli.

2. Proporsi Kumulatif Varians

Proporsi kumulatif varians diperoleh dengan menjumlahkan proporsi varians dari setiap komponen secara berurutan. Menurut Hair, *et al.* (2010), komponen utama dianggap sudah cukup baik untuk merepresentasikan informasi data asli apabila secara kumulatif mampu menjelaskan minimal 70% hingga 90% dari total varians.

3. *Scree Plot*

Penentuan jumlah komponen dilakukan dengan melihat titik siku (*elbow*), yaitu titik di mana penurunan nilai eigen mulai melandai. Komponen-komponen utama yang berada sebelum titik siku dianggap signifikan karena menjelaskan proporsi varians yang relatif besar, sedangkan komponen setelah titik siku dianggap kurang penting karena hanya menjelaskan varians yang kecil (Hair, *et al.*, 2010).

2.2.2 Koefisien Komponen Utama

Dalam Analisis Komponen Utama, koefisien pembobot (*weight*) adalah nilai-nilai yang digunakan untuk membentuk kombinasi linear dari variabel-variabel asli dalam membangun komponen utama. Menurut Jolliffe (2002), koefisien ini diperoleh dari hasil dekomposisi matriks korelasi dalam bentuk vektor eigen (*eigenvector*) yang sesuai dengan nilai eigen (*eigenvalue*) dari masing-masing komponen utama. Secara matematis, jika suatu data memiliki p variabel, maka diperoleh kombinasi linear komponen utama ke- i (KU_i) sebagai berikut:

$$\begin{aligned}
KU_1 &= a_{11}X_1 + a_{12}X_2 + \cdots + a_{1p}X_p \\
KU_2 &= a_{21}X_1 + a_{22}X_2 + \cdots + a_{2p}X_p \\
&\vdots \quad \quad \quad \vdots \quad \quad \quad \vdots \quad \quad \quad \vdots \\
KU_p &= a_{k1}X_1 + a_{k2}X_2 + \cdots + a_{kp}X_p
\end{aligned} \tag{2.1}$$

dimana:

KU_i : Komponen Utama ke- i

a_{ij} : Vektor eigen dari matriks korelasi yang bersesuaian dengan nilai eigen λ_i

X_j : Variabel asli ke- j dengan $j = 1, 2, \dots, p$

Nilai-nilai pembobot ini menunjukkan seberapa besar pengaruh masing-masing variabel terhadap komponen utama yang bersangkutan. Variabel dengan pembobot yang besar (baik positif maupun negatif) menunjukkan kontribusi dominan dalam pembentukan komponen.

2.3 Analisis Biplot

2.3.1 Definisi Analisis Biplot

Analisis biplot pertama kali diperkenalkan oleh Gabriel pada tahun 1971 sebagai metode grafik untuk menyajikan informasi dari suatu matriks data dalam ruang dua dimensi. Menurut Solimun & Fernandes (2008), analisis biplot merupakan salah satu metode visualisasi data dalam analisis multivariat yang memungkinkan penyajian objek (baris) dan variabel (kolom) secara bersamaan dalam suatu grafik berdimensi rendah (2 atau 3). Menurut Mattjik dan Sumertajaya (2011), tujuan utama analisis biplot adalah untuk menggambarkan secara grafis hubungan antar variabel, antar objek, serta hubungan antara objek dan variabel secara bersamaan.

Menurut Ghozali (2018), biplot digunakan untuk menggambarkan pola data multivariat, terutama untuk mengidentifikasi kelompok objek yang memiliki

karakteristik yang sama serta variabel yang berkontribusi besar terhadap perbedaan antar objek tersebut. Dalam biplot, objek digambarkan sebagai titik, sedangkan variabel digambarkan sebagai vektor. Arah dan panjang vektor menunjukkan kontribusi variabel terhadap komponen utama, sedangkan jarak antar titik objek menggambarkan kemiripan karakteristik objek tersebut.

2.3.2 Penguraian Nilai Singular (*Singular Value Decomposition*)

Menurut Jolliffe (2002), analisis biplot klasik didasarkan pada penguraian nilai singular atau *Singular Value Decomposition* (SVD) yang menyatakan jika matriks \mathbf{X} adalah matriks data berukuran $n \times p$, dengan n adalah jumlah objek pengamatan dan p adalah jumlah variabel, maka matriks tersebut dapat diuraikan sebagai berikut:

$$\mathbf{X} = \mathbf{U}\mathbf{L}\mathbf{A}' \quad (2.2)$$

dimana:

1. Matriks \mathbf{X} didefinisikan sebagai matriks data berukuran ($n \times p$) yang memuat sebanyak n objek pengamatan dan sebanyak p variabel yang diamati. Secara umum, bentuk matriks \mathbf{X} adalah:

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} \quad (2.3)$$

2. Matriks \mathbf{U} merupakan matriks berukuran $n \times r$ dan bersifat *orthonormal* sehingga $\mathbf{U}'\mathbf{U} = \mathbf{I}$, dimana \mathbf{I} adalah matriks identitas berdimensi $r \times r$. Setiap kolom dari matriks \mathbf{U} berisi vektor eigen dari matriks $\mathbf{X}\mathbf{X}'$. Sehingga, matriks \mathbf{U} dapat didefinisikan sebagai:

$$\mathbf{U} = [\mathbf{u}_1 \quad \mathbf{u}_2] \quad (2.4)$$

dimana:

$$\mathbf{u}_1 = \begin{bmatrix} u_{11} \\ u_{12} \\ \vdots \\ u_{1n} \end{bmatrix} \quad \mathbf{u}_2 = \begin{bmatrix} u_{21} \\ u_{22} \\ \vdots \\ u_{2n} \end{bmatrix} \quad (2.5)$$

3. Matriks L merupakan matriks diagonal berukuran $r \times r$. Unsur-unsur diagonal matriks L merupakan akar kuadrat dari nilai eigen matriks $X'X$ atau XX' , sehingga unsur-unsur diagonalnya:

$$\sqrt{\lambda_1} \geq \sqrt{\lambda_2} \geq \dots \geq \sqrt{\lambda_r} \quad (2.6)$$

Matriks L dapat didefinisikan sebagai:

$$L = \begin{bmatrix} \sqrt{\lambda_1} & 0 & \dots & 0 \\ 0 & \sqrt{\lambda_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sqrt{\lambda_r} \end{bmatrix} \quad (2.7)$$

4. Matriks A berukuran $(p \times r)$ dan memiliki kolom *orthonormal* sehingga $A'A = I$, dimana I adalah matriks identitas berdimensi $r \times r$. Setiap kolom matriks A adalah vektor eigen dari matriks $X'X$ yang bersesuaian dengan nilai eigen.

Matriks A dapat didefinisikan sebagai:

$$A = [a_1, a_2, \dots, a_r] \quad (2.8)$$

dimana a_i menyatakan kolom ke- i dari matriks A

$$a_1 = \begin{bmatrix} a_{11} \\ a_{12} \\ \vdots \\ a_{1n} \end{bmatrix} \quad a_2 = \begin{bmatrix} a_{21} \\ a_{22} \\ \vdots \\ a_{2n} \end{bmatrix} \quad (2.9)$$

Dari hasil penguraian nilai singular (SVD) pada matriks data X , dapat dibentuk grafik biplot dengan menentukan matriks G dan H . Menurut Jolliffe (2002), misalkan matriks $G = UL^\alpha$ dan $H = LA^\alpha$ atau $H' = L^{1-\alpha}A'$, maka persamaannya dapat ditulis menjadi:

$$X = UL^\alpha L^{1-\alpha}A' = GH' \quad (2.10)$$

dengan nilai α pada rentang $0 \leq \alpha \leq 1$. Nilai α digunakan untuk mengontrol keseimbangan representasi antara objek (baris) dan variabel (kolom) dalam grafik biplot.

1. Untuk $\alpha = 1$, representasi fokus pada objek (koordinat objek dipertahankan penuh, sedangkan variabel diturunkan dari objek).
2. Untuk $\alpha = 0$, representasi fokus pada variabel (vektor variabel dipertahankan penuh, sedangkan objek diturunkan dari variabel).

3. Untuk $\alpha = 0.5$: keseimbangan antara keduanya (memberikan bobot setara antara objek dan variabel sehingga hasil visualisasi menjadi seimbang).

Dari pendekatan matriks \mathbf{X} pada dimensi dua diperoleh matriks \mathbf{G} dan \mathbf{H} sebagai berikut:

$$\mathbf{G} = \begin{bmatrix} g_{11} & g_{12} \\ \vdots & \vdots \\ g_{i1} & g_{i2} \\ \vdots & \vdots \\ g_{n1} & g_{n2} \end{bmatrix} \quad \text{dan} \quad \mathbf{H} = \begin{bmatrix} h_{11} & h_{12} \\ \vdots & \vdots \\ h_{i1} & h_{i2} \\ \vdots & \vdots \\ h_{p1} & h_{p2} \end{bmatrix} \quad (2.11)$$

dimana matriks \mathbf{G} berukuran $n \times r$ merupakan komponen utama yang mencerminkan koordinat titik dari n objek, sementara matriks \mathbf{H} berukuran $p \times r$ merupakan vektor eigen yang mencerminkan koordinat vektor dari p variabel. Dengan menggunakan dua kolom pertama dari masing-masing matriks, maka matriks data \mathbf{X} dapat ditunjukkan dalam bentuk biplot.

2.3.3 Interpretasi Biplot

Menurut Susetyo (2003), ada empat informasi penting yang bisa diperoleh dari analisis biplot, yaitu:

1. Kedekatan antar objek yang diamati

Kedekatan antar objek mengindikasikan tingkat kemiripan karakteristik di antara objek-objek yang diamati. Objek yang berada berdekatan satu sama lain dalam ruang biplot cenderung memiliki profil atau nilai variabel yang serupa. Sebaliknya, objek yang berjauhan menunjukkan perbedaan karakteristik yang lebih besar.

2. Keragaman variabel

Keragaman variabel menunjukkan sejauh mana nilai suatu variabel bervariasi antar objek yang diamati. Variabel dengan nilai yang sangat bervariasi menunjukkan keragaman tinggi, sedangkan variabel dengan nilai yang hampir seragam menunjukkan keragaman rendah. Dalam analisis biplot, keragaman

variabel direpresentasikan oleh panjang vektor masing-masing variabel. Vektor yang panjang menunjukkan keragaman tinggi, sedangkan vektor yang pendek menunjukkan keragaman rendah.

3. Korelasi antar variabel

Dalam analisis biplot, korelasi antar variabel ditunjukkan oleh sudut antar vektor yang mewakili masing-masing variabel. Dua variabel dikatakan memiliki korelasi positif apabila dua buah vektor yang membentuk sudut lancip ($< 90^\circ$). Dua variabel dikatakan memiliki korelasi negatif apabila dua buah vektor yang membentuk sudut tumpul ($> 90^\circ$). Sedangkan dua variabel dikatakan tidak memiliki korelasi apabila dua buah vektor yang membentuk sudut siku-siku (90°).

4. Nilai variabel pada suatu objek

Nilai suatu variabel pada objek tertentu menggambarkan tingkat keunggulan atau kelemahan objek terhadap variabel tersebut. Dalam biplot, objek yang terletak searah dengan arah vektor variabel (yang digambarkan sebagai garis atau panah) menunjukkan bahwa objek tersebut memiliki nilai lebih tinggi dibandingkan objek lainnya pada variabel tersebut. Sebaliknya, jika suatu objek terletak berlawanan arah dengan vektor variabel, maka objek tersebut memiliki nilai lebih rendah pada variabel tersebut dibandingkan objek lainnya.

2.3.4 Ukuran Kesesuaian Biplot

Dalam analisis biplot, ukuran kesesuaian (*goodness of fit*) mengacu pada sejauh mana grafik biplot mampu merepresentasikan informasi dari data asli ke dalam ruang berdimensi rendah, biasanya dua dimensi. Menurut Rencher (2002), ukuran kesesuaian biplot dapat dievaluasi dengan menguji dua nilai eigen λ_1 dan λ_2 dengan pendekatan matriks \mathbf{X} berdimensi dua dalam bentuk:

$$\rho^2 = \frac{\lambda_1 + \lambda_2}{\sum_{k=1}^m \lambda_k} \quad (2.12)$$

dimana:

ρ^2 : ukuran ketepatan biplot dua dimensi

λ_1 : nilai eigen terbesar ke-1

λ_2 : nilai eigen terbesar ke-2

λ_k : total nilai eigen keseluruhan

Menurut Mattjik dan Sumertajaya (2011), apabila ρ^2 mendekati nilai satu ($\geq 70\%$), maka biplot memberikan penyajian yang semakin baik mengenai informasi data yang sebenarnya.

2.4 *K-Means Cluster*

Menurut MacQueen (1967), *K-Means* merupakan salah satu metode klasterisasi non-hierarki yang digunakan untuk mengelompokkan data ke dalam sejumlah klaster berdasarkan kemiripan karakteristik yang dimiliki. Metode ini bekerja dengan menempatkan objek pada klaster yang memiliki pusat klaster (*centroid*) terdekat, sehingga objek dalam kelompok yang sama memiliki karakteristik yang homogen, sedangkan objek antar kelompok memiliki karakteristik yang heterogen. Menurut Jain (2010), *K-Means* termasuk metode *partitional clustering* yang bertujuan meminimalkan variasi di dalam klaster dan memaksimalkan perbedaan antar klaster. Berikut adalah langkah-langkah dalam metode *K-Means*:

1. Menentukan terlebih dahulu banyaknya klaster yang akan dibentuk. Misalnya ($k = 2$).
2. Menentukan *centroid* awal untuk setiap klaster, yang umumnya dipilih secara acak dari data. Misalnya:

$$\begin{aligned} C_1 &= (x_1, y_1) \\ C_2 &= (x_2, y_2) \end{aligned} \quad (2.13)$$

3. Menghitung jarak antara setiap objek dengan seluruh *centroid* menggunakan jarak *Euclidean*. Jarak *Euclidean* antara objek i dan j dirumuskan dengan:

$$d(x, y) = \sqrt{\sum_{k=1}^n (x_{ik} - x_{jk})^2} \quad (2.14)$$

dimana:

d_{ij} = jarak objek ke- i terhadap *centroid* ke- j ,

x_{ik} = nilai variabel ke- k pada objek ke- i ,

x_{jk} = nilai variabel ke- k pada *centroid* ke- j ,

n = jumlah variabel.

4. Mengelompokan objek berdasarkan jarak terdekat antara setiap objek terhadap *centroid*.
5. Menentukan *centroid* baru dengan menghitung rata-rata seluruh objek pada pusat kluster yang sama.

$$C_{kj} = \frac{x_{1j} + x_{2j} + \dots + x_{aj}}{a}, \quad j = 1, 2, 3, \dots, p \quad (2.15)$$

C_{kj} = *centroid* ke- j ,

a = jumlah anggota kluster ke- j .

6. Ulangi langkah 3 sampai 5 hingga tidak terjadi perubahan keanggotaan kluster atau perubahan *centroid* sudah sangat kecil (*konvergen*).
7. Setelah diperoleh kluster yang stabil, dilakukan evaluasi kualitas kluster menggunakan ukuran validasi, seperti:
 - a. *Silhouette Coefficient*
 - b. *Elbow Method (SSE/WCSS)*
 - c. *Calinski-Harabasz Index*
 - d. *Davies-Bouldin Index*

2.5 Standarisasi Data

Menurut Kuncoro (2003), standarisasi data adalah proses mengubah nilai suatu data agar berada pada skala yang sama. Jika variabel memiliki skala yang tidak seragam, maka variabel dengan nilai yang lebih besar akan memberikan pengaruh yang lebih dominan terhadap hasil analisis dibandingkan dengan variabel yang memiliki nilai kecil. Untuk mengatasi perbedaan tersebut, perlu dilakukan proses standarisasi data agar masing-masing variabel memiliki pengaruh yang setara.

Salah satu metode yang umum digunakan adalah *Z-Score Standardization*, yaitu metode yang mengubah data sehingga memiliki rata-rata (*mean*) sebesar 0 dan simpangan baku (*standar deviasi*) sebesar 1. Menurut Larose (2005), rumus standarisasi data dengan metode *Z-score* dapat dituliskan sebagai berikut:

$$Z = \frac{X_i - \bar{X}}{s}, \quad i = 1, 2, \dots, p \quad (2.16)$$

dimana:

Z : nilai *Z-Score*

X_i : nilai data asli ke- i

\bar{X} : rata-rata sampel

s : simpangan baku sampel

2.6 Kesejahteraan Masyarakat

Kesejahteraan masyarakat merupakan suatu kondisi yang menggambarkan terpenuhinya kebutuhan dasar masyarakat secara layak, baik dari segi ekonomi, kesehatan, pendidikan, maupun kualitas hidup secara menyeluruh. Menurut Suparlan (2004), kesejahteraan masyarakat adalah kondisi terpenuhinya kebutuhan material dan spiritual masyarakat, yang mencakup kebutuhan sandang, pangan, papan, pendidikan, kesehatan, dan rasa aman. Dengan demikian, kesejahteraan tidak hanya dilihat dari aspek ekonomi, tetapi juga dari aspek sosial dan psikologis.

Menurut Badan Pusat Statistik Indonesia (2024), kesejahteraan masyarakat di Indonesia dinilai berdasarkan berbagai indikator multidimensi yang mencakup beberapa aspek, yaitu:

1. Kependudukan; meliputi jumlah penduduk, laju pertumbuhan penduduk, kepadatan penduduk, dan struktur penduduk.
2. Kesehatan dan Gizi; diukur melalui umur harapan hidup, akses terhadap fasilitas kesehatan, status kesehatan, serta kondisi gizi masyarakat.
3. Pendidikan; mencakup rata-rata lama sekolah, angka melek huruf, tingkat pendidikan yang ditamatkan dan indikator pendidikan lainnya.

4. Ketenagakerjaan; yang meliputi tingkat partisipasi angkatan kerja, tingkat pengangguran terbuka, dan kesempatan kerja.
5. Taraf dan Pola Konsumsi; menggambarkan kondisi ekonomi masyarakat melalui pengeluaran per kapita dan pola konsumsi rumah tangga.
6. Perumahan dan Lingkungan; mencakup kondisi fisik rumah, akses terhadap air bersih, sanitasi, dan fasilitas dasar lainnya.
7. Kemiskinan; diukur melalui tingkat kemiskinan, indeks kedalaman kemiskinan, dan rasio gini.
8. Indikator Sosial Lainnya; meliputi partisipasi sosial masyarakat, tingkat kriminalitas, akses terhadap informasi & teknologi, serta berbagai aspek sosial yang memengaruhi kualitas hidup dan kesejahteraan masyarakat.

Penelitian ini menggunakan beberapa variabel kesejahteraan masyarakat sebagaimana ditetapkan oleh Badan Pusat Statistik Indonesia (2024). Adapun variabel-variabel tersebut sebagai berikut:

1. Umur Harapan Hidup

Rata-rata perkiraan banyak tahun yang dapat ditempuh oleh seseorang sejak lahir apabila pola mortalitas yang berlaku pada saat tertentu tetap berlaku sepanjang hidupnya.

2. Jumlah Rumah Sakit

Banyaknya rumah sakit yang beroperasi pada suatu wilayah dan memberikan pelayanan kesehatan kepada masyarakat. Rumah sakit merupakan institusi pelayanan kesehatan yang menyelenggarakan pelayanan rawat inap, rawat jalan, dan gawat darurat.

3. Rata-rata Lama Sekolah

Rata-rata jumlah tahun yang digunakan oleh penduduk usia 25 tahun ke atas dalam menjalani pendidikan formal. RLS digunakan untuk menggambarkan tingkat pendidikan yang telah dicapai masyarakat.

4. Presentase Penduduk Tamat SMA

Persentase penduduk yang telah menyelesaikan pendidikan pada jenjang Sekolah Menengah Atas (SMA), Sekolah Menengah Kejuruan (SMK),

Madrasah Aliyah (MA), atau pendidikan sederajat. Indikator ini berasal dari karakteristik jenjang pendidikan tertinggi yang ditamatkan

5. Tingkat Pengangguran Terbuka

Pengangguran terbuka mencakup penduduk yang tidak bekerja dan sedang mencari pekerjaan, mempersiapkan usaha, atau menunggu panggilan kerja.

$$TPT = \frac{\text{Jumlah Pengangguran}}{\text{Jumlah Angkatan Kerja}} \times 100\% \quad (2.17)$$

6. Tingkat Partipasi Angkatan Kerja

Angkatan kerja terdiri atas penduduk yang bekerja dan pengangguran yang sedang mencari pekerjaan.

$$TPAK = \frac{\text{Jumlah Angkatan Kerja}}{\text{Penduduk Usia Kerja}} \times 100\% \quad (2.18)$$

7. Pengeluaran per Kapita Disesuaikan

Rata-rata pengeluaran konsumsi penduduk per orang per tahun yang telah disesuaikan dengan daya beli masyarakat. Pengeluaran per kapita mencerminkan kemampuan masyarakat dalam memenuhi kebutuhan hidupnya.

8. Hunian Layak

Persentase rumah tangga yang menempati hunian yang memenuhi kriteria kelayakan, meliputi kualitas bangunan yang memadai, kecukupan luas lantai minimal 7,2 m² per kapita, memiliki akses terhadap sumber air minum layak, serta memiliki akses sanitasi layak

9. Gini Ratio

Ukuran ketimpangan distribusi pendapatan atau pengeluaran dalam suatu wilayah dengan nilai antara 0 dan 1. Nilai mendekati 0 menunjukkan pemerataan yang lebih baik, sedangkan nilai mendekati 1 menunjukkan ketimpangan yang lebih tinggi.

10. Korban Kejahatan

Persentase atau jumlah penduduk yang pernah mengalami tindak kejahatan dalam periode tertentu. Tindak kejahatan dapat berupa pencurian, penganiayaan, perampokan, pelecehan seksual, atau tindak kriminal lainn

III. METODOLOGI PENELITIAN

3.1 Waktu dan Tempat Penelitian

Penelitian ini dilaksanakan pada semester genap tahun ajaran 2025/2026, bertempat di Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.

3.2 Data Penelitian

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari *website* Badan Pusat Statistik (BPS) Provinsi Lampung dengan situs: <https://lampung.bps.go.id/id>. Data yang digunakan berupa 10 variabel indikator kesejahteraan masyarakat pada 15 kab/kota di Provinsi Lampung pada tahun 2025. Berikut ini adalah variabel-variabel yang digunakan dalam penelitian:

Tabel 1. Variabel Indikator Kesejahteraan Masyarakat

Variabel	Keterangan	Satuan
X_1	Umur Harapan Hidup	Tahun
X_2	Jumlah Rumah Sakit	Unit
X_3	Rata-rata Lama Sekolah	Tahun
X_4	Persentase Penduduk Tamat SMA	%
X_5	Tingkat Pengangguran Terbuka	%
X_6	Tingkat Partisipasi Angkatan Kerja	%
X_7	Pengeluaran per Kapita disesuaikan	Rp/orang/tahun

Tabel 1. Lanjutan

Variabel	Keterangan	Satuan
X_8	Hunian Layak	%
X_9	Gini Ratio	Indeks
X_{10}	Korban Kejahatan	%

3.3 Metode Penelitian

Metode yang digunakan dalam penelitian ini adalah *Principal Component Analysis* (PCA) Biplot dan *K-Means* yang diolah menggunakan bantuan *software R-Studio*. Adapun langkah- langkah yang dilakukan adalah sebagai berikut:

1. Mendeskripsikan data menggunakan analisis deskriptif.
2. Melakukan standarisasi data.
3. Melakukan *Principal Component Analysis* (PCA), yang meliputi:
 - a) Menentukan komponen utama berdasarkan nilai eigen ≥ 1 , proporsi kumulatif varians sebesar 70% - 90%, dan *scree plot*.
 - b) Menentukan nilai koefisien komponen utama terpilih (KU_{1i} dan KU_{2i})
 - c) Menghitung skor komponen utama melalui transformasi data ke dalam ruang komponen utama.
4. Melakukan analisis biplot, yang meliputi:
 - a) Membentuk grafik biplot berdasarkan vektor (v_{1j}, v_{2j}) sebagai representasi variabel X_j dan titik koordinat (KU_{1i}, KU_{2i}) sebagai representasi objek i ;
 - b) Menginterpretasikan grafik biplot.
 - c) Menghitung ukuran kesesuaian biplot (*goodness of fit*) berdasarkan proporsi dua nilai eigen terbesar pertama.
5. Melakukan analisis *K-Means*, yang meliputi:
 - a) Menentukan jumlah kluster optimum.
 - b) Membentuk kluster optimum.
 - c) Menginterpretasikan karakteristik masing-masing kluster.

V. KESIMPULAN

Berdasarkan hasil analisis dan pembahasan yang telah dilakukan sebelumnya, maka diperoleh beberapa kesimpulan sebagai berikut:

1. Hasil *Principal Component Analysis* (PCA) menunjukkan bahwa terbentuk dua komponen utama yang mampu menjelaskan 71,22% keragaman data. KU_1 didominasi oleh variabel Umur Harapan Hidup, Jumlah Rumah Sakit, Rata-Rata Lama Sekolah, Presentase Penduduk Tamat SMA, Pengeluaran per Kapita dan Gini Ratio dengan nilai keragaman sebesar 54,32%. Sementara itu, KU_2 didominasi oleh variabel Tingkat Pengangguran Terbuka, Tingkat Partisipasi Angkatan Kerja, Hunian Layak, dan Korban Kejahatan dengan nilai keragaman sebesar 16,89%.
2. Hasil analisis *K-Means*, menunjukkan bahwa kabupaten/kota di Provinsi Lampung terbagi ke dalam dua klaster. Klaster 1 terdiri atas 13 kabupaten yang memiliki nilai KU_1 relatif rendah dan KU_2 relatif tinggi, sedangkan Klaster 2 terdiri atas Kota Bandar Lampung dan Kota Metro yang memiliki nilai KU_1 relatif tinggi dan KU_2 relatif rendah.
3. Klaster 1 merupakan kelompok wilayah dengan tingkat kesejahteraan yang relatif lebih rendah, namun memiliki kondisi ketenagakerjaan yang lebih baik serta tingkat ketimpangan pendapatan dan jumlah korban kejahatan yang lebih rendah. Sebaliknya, klaster 2 merupakan kelompok wilayah dengan tingkat kesejahteraan yang relatif lebih tinggi, tetapi menghadapi tingkat pengangguran, ketimpangan pendapatan, dan jumlah korban kejahatan yang lebih besar dibandingkan klaster 1.

DAFTAR PUSTAKA

- Badan Pusat Statistik Indonesia. 2024. *Indikator Kesejahteraan Rakyat Tahun 2024*. BPS, Jakarta.
- Badan Pusat Statistik Indonesia. 2023. *Indeks Pembangunan Manusia 2023*. BPS, Jakarta.
- Badan Pusat Statistik Provinsi Lampung. 2025. *Data Indikator Kesejahteraan Masyarakat*. <https://lampung.bps.go.id/id> diakses pada Januari 2026.
- Badan Pusat Statistik Provinsi Lampung. 2023. *Statistik daerah Provinsi Lampung 2023*. BPS Provinsi Lampung, Bandar Lampung.
- Ghozali, I. 2018. *Aplikasi Analisis Multivariate dengan Program IBM SPSS 25 (Edisi 9)*. Badan Penerbit Universitas Diponegoro, Semarang.
- Hair, J. F., Anderson, R. E., Tatham, R. I., & Black, W. C.. 2010. *Multivariate data analysis*. 7th edition. Prentice Hall, New Jersey.
- Hutagalung, R. & Nuramaliyah. 2026. Pengelompokan Kabupaten/Kota di Sumatera Utara Berdasarkan Indikator Infrastruktur dan Layanan Dasar Menggunakan *Principal Component Analysis* dan *K-Means Clustering*. *Prosiding Seminar Nasional Sains dan Teknologi "SainTek"*. 3(1).
- Jain, A. K. 2010. Data Clustering: 50 Years Beyond K-Means. *Pattern Recognition Letters*. 31(8): 651–666.
- Johnson, R.A. & Wichern, D.W. 2007. *Applied Multivariate Statistical Analysis*. 6th edition. Prentice Hall, New Jersey.

- Jolliffe, I. T. 2002. *Principal Component Analysis Second Edition*. Springer Verlag, New York.
- Kuncoro, M. 2003. *Metode kuantitatif: Teori dan aplikasi untuk bisnis dan ekonomi*. UPP AMP YKPN, Yogyakarta.
- Larose, D. T. 2005. *Discovering statistics*. NJ: John Wiley & Sons, Hoboken.
- MacQueen, J. 1967. Some Methods for Classification and Analysis of Multivariate Observations. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1*. University of California Press, Berkeley. 281–297.
- Mattjik, A. A. & Sumertajaya, I. M. 2011. *Sidik peubah ganda*. IPB Press, Bogor.
- Rencher, A. C. 2002. *Methods of Multivariate Analysis Second Edition*. A Jhon Wiley & Sons, Inc., Amerika.
- Rosyada, I. A. & Utari, D. T. 2024. Penerapan *Principal Component Analysis* untuk Reduksi Variabel pada Algoritma *K-Means Clustering*. *Jambura Journal of Probability and Statistics*. 5(1): 6–13.
- Solimun & Fernandes, A. A. R. 2008. *Modul Pelatihan Multivariate Analysis 6&8 Juni 2008*. LPM UB, Malang.
- Suparlan, P. 2004. *Kemiskinan di Perkotaan*. Yayasan Obor Indonesia, Jakarta.
- Susetyo, B. 2003. *Statistika Multivariat: Konsep dan Aplikasinya dengan SPSS*. PT Gramedia Pustaka Utama, Jakarta.
- Widyatama, H. 2017. Clustering Wilayah Lampung Berdasarkan Tingkat Kesejahteraan. *Prosiding Seminar Nasional Metode Kuantitatif*. 149–155.