

Analisis Komponen Utama dan Klasterisasi K-Means pada Data Pengangguran Terbuka Kabupaten/Kota di Provinsi Sulawesi Selatan

Principal Component Analysis and K-Means Clustering on Open Unemployment Data of Regencies/Cities in South Sulawesi Province

Muhammad Hidayatullah^{*1}, Afif Budi Andy B², Irwan Usman³

Universitas Sulawesi Barat^{1,2,3}

E-mail : muh.hidayatullah@unsulbar.ac.id ^{*1}, afifbudiandy.b@unsulbar.ac.id ²,
irwan.usman@unsulbar.ac.id ³

**Corresponding author*

Received 3 May 2026; Revised 14 May 2026; Accepted 15 May 2026

Abstrak - Pengangguran terbuka merupakan salah satu permasalahan ketenagakerjaan yang masih menjadi tantangan di Provinsi Sulawesi Selatan, dengan tingkat disparitas yang tinggi antar kabupaten/kota. Penelitian ini bertujuan untuk mengelompokkan 24 kabupaten/kota di Provinsi Sulawesi Selatan berdasarkan pola Tingkat Pengangguran Terbuka (TPT) tahun 2019–2023 menggunakan kombinasi metode *Principal Component Analysis* (PCA) dan *K-Means clustering*. Data yang digunakan merupakan data sekunder TPT hasil Survei Angkatan Kerja Nasional (Sakernas) periode Agustus yang bersumber dari Badan Pusat Statistik (BPS) Provinsi Sulawesi Selatan. Analisis PCA dilakukan untuk mereduksi dimensi data dari lima variabel tahun pengamatan menjadi beberapa komponen utama yang saling ortogonal, kemudian skor komponen utama digunakan sebagai input dalam *K-Means clustering*. Hasil penelitian menunjukkan bahwa 24 kabupaten/kota terbagi ke dalam tiga *cluster*, yaitu *Cluster 1* dengan 13 kabupaten/kota berkategori TPT rendah, *Cluster 2* dengan 9 kabupaten/kota berkategori TPT menengah, dan *Cluster 3* dengan 2 wilayah yaitu Kota Makassar dan Kota Palopo yang berkategori TPT tinggi. Hasil pengelompokan ini diharapkan dapat menjadi referensi bagi pemerintah daerah dalam merumuskan kebijakan ketenagakerjaan yang tepat sasaran sesuai karakteristik masing-masing wilayah.

Kata Kunci – *Clustering, K-Means, Principal Component Analysis, Pengangguran Terbuka, Sulawesi Selatan*

Abstract - *Open unemployment remains one of the labor market challenges in South Sulawesi Province, characterized by high levels of disparity across districts and cities. This study aims to classify 24 districts and cities in South Sulawesi Province based on the Open Unemployment Rate (OUR) patterns from 2019 to 2023 using a combination of Principal Component Analysis (PCA) and K-Means clustering methods. The data used are secondary data on the Open Unemployment Rate obtained from the National Labor Force Survey (Sakernas) for the August period, sourced from the Statistics Indonesia (BPS) of South Sulawesi Province. PCA was applied to reduce the dimensionality of the data from five annual observation variables into several mutually orthogonal principal components, after which the principal component scores were used as input for K-Means clustering. The results show that the 24 districts and cities are grouped into three clusters: Cluster 1 consisting of 13 districts and cities with low unemployment rates, Cluster 2 consisting of 9 districts and cities with moderate unemployment rates, and Cluster 3 consisting of 2 areas, namely Makassar City and Palopo City, with high unemployment rates. These findings are expected to serve as a reference for local governments in formulating targeted employment policies tailored to the characteristics of each regional group.*

Keywords – *Clustering, K-Means, Principal Component Analysis, Open Unemployment Rate, Sulawesi Selatan*

1. PENDAHULUAN

Pengangguran merupakan salah satu permasalahan ketenagakerjaan yang menjadi perhatian utama dalam pembangunan ekonomi di Indonesia. Provinsi Sulawesi Selatan merupakan provinsi yang telah mencatat pertumbuhan ekonomi yang cukup signifikan, namun diiringi dengan tantangan pengangguran. Data dari Badan Pusat Statistik (BPS) menunjukkan bahwa tingkat pengangguran terbuka (TPT) tetap menjadi tantangan utama [1]. Kondisi ini semakin kompleks mengingat Sulawesi Selatan memiliki 24 kabupaten/kota dengan karakteristik wilayah, struktur ekonomi, dan kapasitas penyerapan tenaga kerja yang berbeda-beda satu sama lain [2].

Terdapat hubungan timbal balik yang kompleks di antara kabupaten/kota di Provinsi Sulawesi Selatan. Tingkat pengangguran terbuka, indeks pembangunan manusia, dan jumlah penduduk terbukti memiliki dampak yang signifikan terhadap ketimpangan distribusi pendapatan di antara kabupaten dan kota tersebut. Hal ini menunjukkan bahwa persoalan pengangguran tidak berdiri sendiri, melainkan berkaitan erat dengan ketimpangan pembangunan antar wilayah. Oleh karena itu, diperlukan pemetaan yang sistematis untuk memahami pola persebaran pengangguran terbuka di tingkat kabupaten/kota guna mendukung perumusan kebijakan ketenagakerjaan yang lebih tepat sasaran [3].

Periode 2019–2023 merupakan rentang waktu strategis untuk dianalisis karena mencakup kondisi sebelum, selama, dan setelah kemunculan serta penyebaran selanjutnya dari virus korona baru, yang secara resmi dikenal sebagai SARS-CoV-2, yaitu virus penyebab penyakit yang dikenal sebagai *Covid-19*. Pandemi ini berdampak buruk pada sektor-sektor utama perekonomian, dan akibatnya, individu yang siap dan bersedia bekerja tidak dapat memperoleh pekerjaan, sehingga menyebabkan pengangguran jangka pendek hingga menengah yang berkepanjangan. Fenomena ini disebabkan oleh fakta bahwa mekanisme pasar saja tidak cukup untuk mengoptimalkan penawaran dan permintaan tenaga kerja. Fenomena ini menciptakan fluktuasi TPT yang signifikan dan berbeda intensitasnya di setiap kabupaten/kota, sehingga memunculkan kebutuhan untuk mengelompokkan wilayah berdasarkan pola TPT dari tahun ke tahun [2].

Dalam konteks analisis data multivariat dengan dimensi waktu, penggunaan metode statistik lanjutan menjadi sangat relevan. *Principal Component Analysis* (PCA) merupakan langkah prapemrosesan yang sangat penting dalam algoritma pengelompokan *K-Means* untuk mengurangi jumlah variabel. Analisis Komponen Utama (PCA) adalah teknik statistik yang mengidentifikasi variabel-variabel yang paling berkontribusi terhadap varians total dalam data. PCA memungkinkan penghilangan variabel-variabel yang berlebihan dalam analisis dan memungkinkan algoritma *K-Means* untuk berfokus pada fitur-fitur yang paling informatif [4].

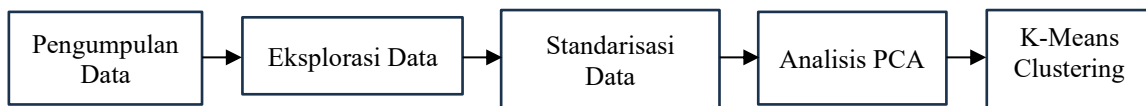
Pengelompokan dari analisis *cluster* dengan metode *K-Means* yang dikombinasikan dengan PCA terbukti mampu menghasilkan kelompok-kelompok yang lebih optimal dibandingkan *K-Means* tanpa reduksi dimensi [4][5]. Pendekatan ini relevan untuk diterapkan pada data TPT kabupaten/kota di Sulawesi Selatan yang memiliki variabel pengamatan lintas tahun (2019–2023), di mana setiap tahun diperlakukan sebagai variabel sehingga berpotensi mengandung multikolinearitas dan redundansi informasi.

Beberapa penelitian sebelumnya telah menerapkan *K-Means clustering* untuk mengelompokkan wilayah berdasarkan indikator ketenagakerjaan. Penelitian [6] melakukan *K-Means clustering* pada persebaran tingkat pengangguran kabupaten/kota di Sulawesi Selatan dengan menggunakan indikator upah minimum kabupaten (UMK) dan laju pertumbuhan Indeks Pembangunan Manusia (IPM), dan berhasil mengelompokkan 24 kabupaten/kota ke dalam beberapa kluster yang merepresentasikan karakteristik pengangguran di wilayah tersebut. Namun, penelitian tersebut belum mengintegrasikan analisis PCA sebagai tahap reduksi dimensi sebelum proses klusterisasi, serta belum mempertimbangkan dinamika perubahan TPT lintas waktu yang mencakup periode pandemi.

Tantangan utama yang dihadapi dalam pembangunan daerah adalah belum adanya sistem pengelompokan yang komprehensif untuk kabupaten dan kota yang dapat secara memadai menggambarkan karakteristik sosial-ekonomi mereka. Berdasarkan hal tersebut, penelitian ini bertujuan untuk melakukan analisis komponen utama (PCA) terhadap data TPT kabupaten/kota di Sulawesi Selatan selama periode 2019–2023, kemudian menggunakan skor komponen utama yang dihasilkan sebagai input dalam klasterisasi *K-Means*. Pendekatan ini diharapkan dapat menghasilkan pengelompokan wilayah yang lebih akurat dan informatif, serta memberikan gambaran pola pengangguran yang dapat dijadikan dasar rekomendasi kebijakan ketenagakerjaan daerah.

2. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan menerapkan teknik data mining untuk mengelompokkan wilayah di Sulawesi Selatan berdasarkan tingkat pengangguran terbuka. Adapun alur penelitian dapat dilihat pada Gambar 1.



Gambar 1. Alur Penelitian

Gambar 1 menjelaskan tahapan yang dilakukan yaitu mulai dari pengumpulan data, praproses data, analisis PCA, *clustering*, dan evaluasi hasil *clustering*.

2.1. Pengumpulan Data

Penelitian ini menggunakan data sekunder yang diperoleh dari website Badan Pusat Statistika Sulawesi Selatan (<https://www.bps.go.id>). Data yang digunakan adalah Tingkat Pengangguran Terbuka (TPT) seluruh kabupaten/kota di Provinsi Sulawesi Selatan selama periode tahun 2019 sampai 2023.

2.2. Eksplorasi Data

Eksplorasi data dilakukan untuk mengetahui karakteristik umum data pengangguran terbuka yang digunakan dalam penelitian. Tahapan ini bertujuan untuk memahami distribusi data, nilai rata-rata, penyebaran data, serta mendeteksi kemungkinan adanya perbedaan karakteristik antarvariabel sebelum dilakukan proses *clustering*. Sebelum dilakukan analisis lebih lanjut, dilakukan praproses meliputi eksplorasi data melalui statistik deskriptif untuk memperoleh gambaran umum distribusi TPT pada setiap tahun pengamatan serta melakukan standarisasi data.

2.3. Standarisasi Data

Tahapan praproses terakhir sebelum analisis PCA adalah standarisasi data. Meskipun seluruh variable dalam penelitian ini memiliki satuan yang sama (persen), standarisasi tetap dilakukan untuk memastikan bahwa setiap variable tahun memiliki kontribusi yang proporsional dalam pembentukan komponen utama. Standarisasi dilakukan menggunakan transformasi Z-Score sebagai berikut [7]:

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j} \quad (1)$$

Di mana :

z_{ij} = Nilai TPT yang telah distandarisasi untuk kab/kota ke-i pada tahun ke-j

x_{ij} = Nilai TPT asli kab/kota ke-i pada tahun ke-j

\bar{x}_j = rata-rata TPT seluruh kab/kota pada tahun ke-j

s_j = simpangan baku TPT pada tahun ke-j

2.4. Analisis PCA

Principal Component Analysis (PCA) adalah teknik statistik yang mengubah sekumpulan besar variabel asli, yang saling berkorelasi tinggi, menjadi sekumpulan variabel baru yang lebih kecil dan saling independen [4][8]. Adapun langkah-langkah yang dilakukan untuk menentukan komponen utama adalah sebagai berikut [12]:

1. Menghitung nilai matriks kovarians untuk menentukan korelasi antar setiap variabel.
2. Menghitung *eigenvalue* menggunakan persamaan

$$|\lambda I - R| = 0 \tag{2}$$

dan *eigenvector* dengan rumus

$$R\vec{v} = \lambda \vec{v} \tag{3}$$

3. Menentukan jumlah komponen utama yang mungkin terbentuk dengan mempertimbangkan *eigenvalue* yang lebih besar atau sama dengan 1.
4. Menghitung transformasi data set baru hasil reduksi dengan PCA dengan persamaan :

$$PC_{at} = \vec{v}_{1a}Z_1 + \vec{v}_{2a}Z_2 + \dots + \vec{v}_{pa}Z_p \tag{4}$$

5. Membentuk komponen matriks korelasi yang menunjukkan besarnya korelasi variabel terhadap skor komponen yang dibentuk menggunakan persamaan :

$$rx_p, PC_t = \vec{v}_{1a}\sqrt{\lambda_t} \tag{5}$$

2.5. K-Means Clustering

K-Means Adalah metode *clustering* berbasis partisi yang mengelompokkan data ke dalam sejumlah *k cluster* berdasarkan jarak terdekat terhadap pusat *cluster* (*centroid*) [9]. Tujuan utama *clustering* adalah membagi kumpulan data menjadi kelompok-kelompok yang relatif homogen, dimana kesamaan antar data dalam satu kelompok dioptimalkan, sedangkan kesamaan dengan data di kelompok lain diminimalkan [10].

Algoritma *K-Means Clustering* dapat dituliskan sebagai berikut [10] :

1. Menentukan jumlah *cluster* (K)
2. Inisialisasi K *centroid* awal secara acak ($C = \{c_1, c_2, \dots, c_K\}$)
3. Ulangi hingga konvergen:
 - a. Hitung jarak setiap data x_i terhadap seluruh *centroid* menggunakan jarak Euclidean.

$$d(x_i, c_j) = \sqrt{\sum_{k=1}^p (x_{ik} - c_{jk})^2} \tag{6}$$

- b. Kelompokkan data x_i ke *cluster* dengan jarak terdekat
 - c. Perbaharui *centroid* setiap *cluster* berdasarkan rata-rata anggota *cluster*.

$$c_j = \frac{1}{n_j} \sum_{i=1}^{n_j} x_i \tag{7}$$

4. Jika *centroid* tidak berubah atau perubahan sangat kecil, maka proses dihentikan.
5. Hasil akhir berubah K *cluster* dan *centroid* akhir.

3. HASIL DAN PEMBAHASAN

3.1. Data Penelitian

Data yang digunakan dalam penelitian ini adalah data tingkat pengangguran terbuka provinsi Sulawesi Selatan yang diperoleh dari website Badan Pusat Statistik (BPS) Provinsi Sulawesi Selatan sebanyak 24 data. Data yang digunakan yaitu data Tingkat pengangguran terbuka 5 tahun terakhir (2019-2023) yang dapat dilihat pada Tabel 1.

Tabel 1. Tingkat Pengangguran Terbuka Kab/Kota di Sulawesi Selatan

Kabupaten/Kota	Tingkat Pengangguran Terbuka Provinsi Sulawesi Selatan				
	2019	2020	2021	2022	2023
Kepulauan Selayar	1,1	2,44	2,81	1,49	1,98
Bulukumba	3,06	3,42	3,14	1,26	1,31

Bantaeng	3,65	4,27	4,07	2,72	3,71
Jeneponto	1,99	2,31	2,38	2,21	2,13
Takalar	3,78	4,16	3,93	2,63	3,58
Gowa	4,35	6,44	4,3	3,26	3,43
Sinjai	2,17	2,65	2,61	1,8	1,69
Maros	4,42	6,28	6,3	5,04	3,64
Pangkep	4,99	5,18	5,86	5,23	5,05
Barru	5,39	6,39	6,74	5,32	5,89
Bone	3,01	3,2	4,15	2,27	2,88
Soppeng	3,24	4,42	3,92	3,4	4,34
Wajo	3	4,33	4,32	2,54	2,1
Sidrap	4,35	5,91	4,93	3,56	3,29
Pinrang	2,91	4,19	4,06	2,79	3,64
Enrekang	2,15	2,44	2,34	0,58	1,53
Luwu	4,38	4,94	4,8	3,85	3,7
Tana Toraja	2,55	2,6	3,09	2,32	3,37
Luwu Utara	2,76	3,01	3,91	2,81	2,63
Luwu Timur	3,81	4,46	4,96	4,48	5,42
Toraja Utara	2,86	3,17	2,61	1,99	2,6
Makassar	9,83	15,92	13,18	11,82	10,6
Pare Pare	6,17	7,14	6,72	5,6	5,86
Palopo	9,67	10,37	8,83	8,2	7,81

3.2. Eksplorasi Data

Statistik deskriptif digunakan untuk memberikan gambaran umum mengenai karakteristik masing-masing variabel tahun. Hasil statistik deskriptif ditunjukkan pada Tabel 2.

Tabel 2. Statistik Deskriptif

Statistik	2019	2020	2021	2022	2023
Mean	3,98	4,99	4,75	3,63	3,84
Std.Dev	2,12	3,00	2,40	2,43	2,13
Min	1,10	2,31	2,34	0,58	1,31
Median	3,45	4,30	4,11	2,80	3,51
Max	9,83	15,92	13,18	11,82	10,60

Berdasarkan Tabel 2, rata-rata TPT tertinggi terjadi pada tahun 2020 sebesar 4,99% dan terendah pada tahun 2022 sebesar 3,63%. Lonjakan rata-rata TPT pada tahun 2020–2021 mengindikasikan adanya dampak pandemi *COVID-19* terhadap kondisi ketenagakerjaan di Sulawesi Selatan, yang kemudian berangsur membaik pada tahun 2022 dan 2023.

Nilai standar deviasi tertinggi terjadi pada tahun 2020 sebesar 3,00, yang menunjukkan bahwa dampak pandemi tidak merata antar kabupaten/kota. Rentang yang sangat lebar antara nilai minimum dan maksimum khususnya pada tahun 2020 dengan rentang sebesar 13,61 poin persentase mengonfirmasi adanya disparitas TPT yang signifikan antar wilayah. Pada seluruh tahun pengamatan, nilai *mean* lebih besar dari *median* yang mengindikasikan bahwa distribusi TPT bersifat condong ke kanan (*right-skewed*), artinya sebagian besar kabupaten/kota memiliki TPT di bawah rata-rata sementara sebagian kecil wilayah memiliki TPT yang sangat tinggi.

Kondisi disparitas yang konsisten ini menjadi landasan pentingnya pengelompokan wilayah melalui analisis kluster pada tahapan berikutnya.

3.2. Standarisasi Data

Sebelum dilakukan analisis PCA, seluruh data TPT kabupaten/kota terlebih dahulu distandarisasi menggunakan transformasi Z-score. Hasil standarisasi data disajikan pada Tabel 3.

Tabel 3 Hasil Standarisasi Data

Kabupaten/Kota	Tingkat Pengangguran Terbuka Provinsi Sulawesi Selatan				
	2019	2020	2021	2022	2023
Kepulauan Selayar	-1,389	-0,867	-0,824	-0,901	-0,892
Bulukumba	-0,445	-0,533	-0,684	-0,997	-1,213
Bantaeng	-0,160	-0,244	-0,288	-0,384	-0,063
Jeneponto	-0,961	-0,911	-1,007	-0,598	-0,820
Takalar	-0,098	-0,281	-0,348	-0,421	-0,125
....
....
Makassar	2,818	3,725	3,586	3,443	3,239
Pare Pare	1,054	0,734	0,838	0,828	0,968
Palopo	2,741	1,834	1,736	1,921	1,902

Hasil standarisasi dari Tabel 3 menunjukkan bahwa sebagian besar kabupaten/kota memiliki nilai Z-Score negatif di seluruh tahun pengamatan, yang berarti TPT mereka berada di bawah rata-rata provinsi. Sebaliknya, Kota Makassar dan Kota Palopo secara konsisten memiliki nilai Z-Score positif dan tinggi di sepanjang periode 2019-2023, mengindikasikan bahwa kedua wilayah tersebut memiliki TPT yang jauh di atas rata-rata provinsi. Data yang telah distandarisasi ini selanjutnya digunakan sebagai input dalam analisis komponen utama (PCA).

3.3. Analisis PCA

Principal Component Analysis (PCA) adalah suatu teknik yang dapat digunakan untuk mereduksi dimensi atau variable tanpa menghilangkan informasi didalamnya. Teknik PCA pada penelitian ini akan mereduksi 5 variabel menjadi 2 variabel (PCA 1 dan PCA 2). Sebelum menginterpretasikan skor komponen, terlebih dahulu dilakukan analisis terhadap nilai *loading vector* (*eigenvector*) untuk memahami kontribusi masing-masing variabel tahun terhadap pembentukan setiap komponen utama. Hasil *loading vector* disajikan pada Tabel 4.

Tabel 4. *Loading Factor* PCA

Variabel	<i>Loading Factor</i> PC1	<i>Loading Factor</i> PC2
TPT 2019 (X ₁)	0,4431	-0,3632
TPT 2020 (X ₂)	0,4468	-0,4949
TPT 2021 (X ₃)	0,4516	-0,1042
TPT 2022 (X ₄)	0,4527	0,2127
TPT 2023 (X ₅)	0,4418	0,7531
Proporsi Varians	95,65%	2,08%

Berdasarkan Tabel 4, PC1 memiliki nilai *loading* positif yang seragam pada seluruh variabel tahun. Keseragaman ini menunjukkan bahwa PC1 tidak membedakan antar periode waktu. Kabupaten/kota yang memiliki skor PC1 tinggi adalah wilayah yang secara konsisten memiliki nilai TPT tinggi di seluruh periode pengamatan, terlepas dari efek pandemi maupun

pemulihan ekonomi. Sementara itu, PC2 memiliki nilai *loading* negatif pada tahun 2019-2021 dan nilai *loading* positif pada tahun 2022-2023. Pola ini mengindikasikan bahwa PC2 merepresentasikan dinamika pemulihan ketenagakerjaan pasca pandemi *covid-19*. PC1 dan PC2 juga menunjukkan bahwa kedua komponen utama mampu menjelaskan dari total varians data yaitu sebanyak 97,72%.

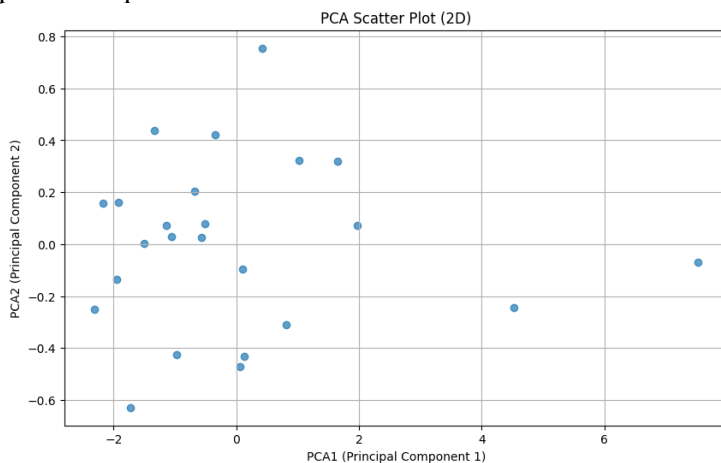
Selanjutnya untuk hasil skor PCA 1 dan PCA 2 untuk setiap data dapat dilihat pada Tabel 5.

Tabel 5. Hasil Analisis PCA

Kabupaten/Kota	PCA	
	PCA 1	PCA 2
Kepulauan Selayar	-2,17699	0,156379
Bulukumba	-1,73151	-0,62888
Bantaeng	-0,5115	0,080036
Jeneponto	-1,92046	0,160103
Takalar	-0,57202	0,027062
....
....
Makassar	7,521819	-0,06872
Pare Pare	1,975799	0,071242
Palopo	4,527847	-0,24309

Berdasarkan hasil skor PCA pada Tabel 5, skor PCA 1 Kota Makassar dan Palopo jauh melampaui seluruh kabupaten/kota lainnya, yang mengonfirmasi bahwa posisi kedua wilayah ini sebagai pencilan atau *outlier* dalam distribusi TPT di Sulawesi Selatan. Sementara itu, ini PCA 2 Kota makassar dan Kota Palopo yang cenderung negatif yang mengindikasikan bahwa meskipun kedua wilayah ini merupakan pusat ekonomi, laju pemulihan TPT pasca-pandemi di kedua wilayah tersebut relatif lebih lambat dibandingkan rata-rata provinsi.

Hasil dari PCA (PCA 1 dan PCA 2) pada Tabel 4 menunjukkan bahwa kedua komponen utama mampu menjelaskan dari total varians data yaitu sebanyak 97,72%. Selain itu, PCA juga digunakan untuk memvisualisasikan pola data dalam ruang berdimensi rendah (2D), sehingga mempermudah melakukan interpretasi *cluster* data. Adapun hasil visualisasi data dari hasil analisis PCA, dapat dilihat pada Gambar 2.



Gambar 2. Visualisasi Data dari Hasil PCA

Berdasarkan visualisasi data pada Gambar 2, menunjukkan bahwa sebagian besar titik data terkonsentrasi di sekitar pusat (titik nol) yang artinya banyak wilayah yang memiliki karakteristik pengangguran tidak terlalu jauh berbeda satu sama lain. Selain itu terdapat beberapa titik data yang jauh dari pusat yang menunjukkan adanya daerah dengan Tingkat pengangguran yang berbeda secara signifikan dari yang lain.

3.3. Clustering K-Means

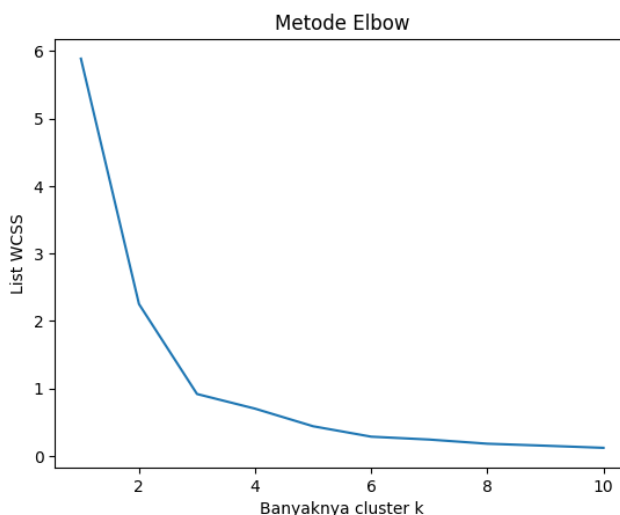
Pada proses ini, dilakukan *clustering* dengan menggunakan algoritma *k-means clustering*. Adapun proses yang dilakukan pada tahap ini sebagai berikut :

a. Menentukan Jumlah *Cluster* (*K*)

Dalam menentukan jumlah *cluster* pada penelitian ini, digunakan metode *Elbow*. Metode ini berfokus pada pemantauan perubahan nilai *Sum of Square Errors* (SSE) atau *Within-Cluster Sum of Square* (WCSS) seiring dengan bertambahnya jumlah *cluster*. Adapun hasil perhitungan SSE / WCSS dapat dilihat pada Tabel 5.

Tabel 6. Hasil Perhitungan WCSS

<i>K</i>	Hasil WCSS
1	5,88
2	2,25
3	0,91
4	0,7
5	0,44
6	0,28
7	0,24
8	0,18
9	0,15
10	0,12



Gambar 3. Visualisasi Metode Elbow

Dari Tabel 6, diperoleh bahwa pada saat nilai $K = 3$, penurunan nilai WCSS mulai melambat secara jelas, membentuk pola seperti elbow seperti pada Gambar 3. Titik ini menandakan bahwa penambahan *cluster* lebih lanjut tidak memberikan pengurangan yang signifikan terhadap WCSS.

b. Menentukan nilai *centroid*

Langkah selanjutnya adalah menentukan nilai *centroid*. Nilai *centroid* dalam penelitian ini ditentukan secara manual, dengan jumlah *centroid* sesuai jumlah *cluster* yang digunakan. Karena pada tahap sebelumnya, diperoleh K optimal = 3, sehingga banyaknya nilai *centroid* yang digunakan juga sebanyak tiga. Adapun nilai *centroid* yang digunakan dapat dilihat pada Tabel 7.

Tabel 7. Titik *Centroid* Awal

<i>Centroid</i>	Nilai <i>Centroid</i>
<i>Centroid 1</i>	[-2.17, 0.15]
<i>Centroid 2</i>	[-1.05, 0.02]
<i>Centroid 3</i>	[4.52, -0.24]

c. Menghitung Jarak *Euclidean*

Jarak antar setiap data terhadap nilai *centroid* dihitung menggunakan *Euclidean Distance*. Berikut contoh perhitungan jarak antar data terhadap nilai *centroid 1*:

$$\text{Data 1} = \sqrt{(-2.17699 - (-2.17))^2 + (0.156379 - 0.15)^2} = 0.00946$$

$$\text{Data 2} = \sqrt{(-1.73151 - (-2.17))^2 + (-0.62888 - 0.15)^2} = 0.893827$$

....

$$\text{Data 24} = \sqrt{(4.527847 - (-2.17))^2 + (-0.24309 - 0.15)^2} = 6.70937$$

Berikut perhitungan jarak antar data terhadap nilai *centroid 2*:

$$\text{Data 1} = \sqrt{(-2.17699 - (-1.05))^2 + (0.156379 - 0.02)^2} = 1.13521$$

$$\text{Data 2} = \sqrt{(-1.73151 - (-1.05))^2 + (-0.62888 - 0.02)^2} = 0.941011$$

....

$$\text{Data 24} = \sqrt{(4.527847 - (-1.05))^2 + (-0.24309 - 0.02)^2} = 5.584048$$

Berikut perhitungan jarak antar data terhadap nilai *centroid 3*:

$$\text{Data 1} = \sqrt{(-2.17699 - 4.52)^2 + (0.156379 - (-0.24))^2} = 6.70871$$

$$\text{Data 2} = \sqrt{(-1.73151 - 4.52)^2 + (-0.62888 - (-0.24))^2} = 6.26359$$

....

$$\text{Data 24} = \sqrt{(4.527847 - 4.52)^2 + (-0.24309 - (-0.24))^2} = 0.00843$$

Berdasarkan perhitungan di atas, hasil perhitungan jarak setiap data terhadap nilai *centroid*, dapat dilihat pada Tabel 7.

Tabel 8. Hasil Perhitungan Jarak Data ke Titik *Centroid*

No	Kabupaten/Kota	Jarak Terpendek
1	Kepulauan Selayar	0.00946 (<i>Centroid 1</i>)
2	Bulukumba	0.89382 (<i>Centroid 1</i>)
...
24	Palopo	0.00843 (<i>Centroid 3</i>)

Iterasi perhitungan jarak terus dilakukan sampai pengelompokan dan nilai *centroid* yang dihasilkan tidak berubah atau nilainya *konvergen*. Dari beberapa iterasi, diperoleh hasil akhir yang dapat dilihat pada Tabel 8.

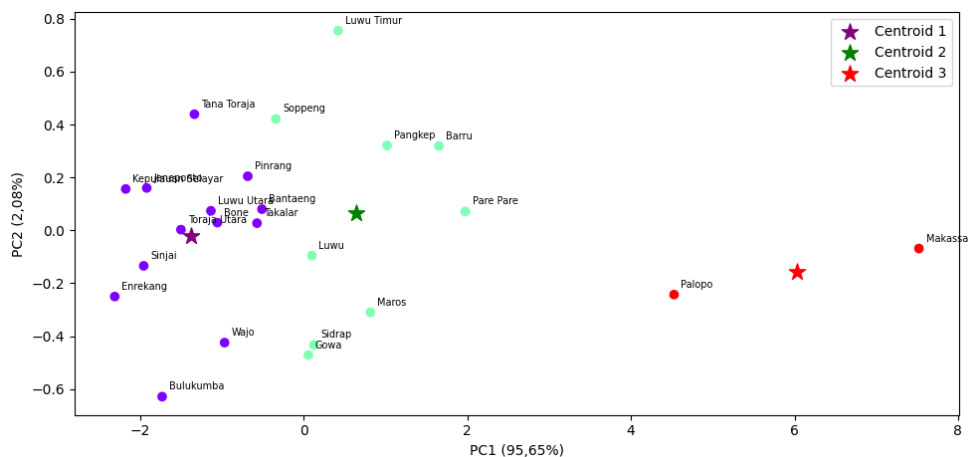
Tabel 9. Hasil Pengelompokan Akhir

No	Kabupaten/Kota	Centroid 1	Centroid 2	Centroid 3	Cluster
1	Kepulauan Selayar	0,8219	2,8248	8,2078	1
2	Bulukumba	0,7055	2,4767	7,7708	1
3	Bantaeng	0,8687	1,1579	6,5406	1
4	Jeneponto	0,5752	2,5685	7,9516	1
5	Takalar	0,8037	1,2189	6,5994	1
6	Gowa	1,4996	0,7970	5,9772	2
7	Sinjai	0,5940	2,6112	7,9822	1
8	Maros	2,2073	0,4100	5,2132	2
9	Pangkep	2,4170	0,4521	5,0291	2
10	Barru	3,0453	1,0375	4,3986	2
11	Bone	0,3211	1,7037	7,0843	1
12	Soppeng	1,1227	1,0510	6,3932	2
13	Wajo	0,5728	1,6869	6,9983	1
14	Sidrap	1,5539	0,7216	5,9076	2
15	Pinrang	0,7250	1,3388	6,7196	1
16	Enrekang	0,9650	2,9745	8,3369	1
17	Luwu	1,4758	0,5696	5,9256	2
18	Tana Toraja	0,4610	2,0183	7,3857	1
19	Luwu Utara	0,2566	1,7819	7,1641	1
20	Luwu Timur	1,9544	0,7272	5,6788	2
21	Toraja Utara	0,1297	2,1491	7,5285	1
22	Makassar	8,8963	6,8768	1,4995	3
23	Pare Pare	3,3514	1,3295	4,0554	2
24	Palopo	5,9064	3,8937	1,4995	3

Berdasarkan hasil pengelompokan akhir pada Tabel 9, diperoleh tiga *cluster* dengan karakteristik yang berbeda. *Cluster 1* beranggotakan 13 kabupaten/kota yaitu Kepulauan Selayar, Bulukumba, Bantaeng, Jeneponto, Takalar, Sinjai, Bone, Wajo, Pinrang, Enrekang, Tana Toraja, Luwu Utara, dan Toraja Utara, yang merupakan wilayah dengan TPT rendah dan stabil, sebagian besar merupakan kabupaten berbasis pertanian di luar kawasan metropolitan. *Cluster 2* beranggotakan 9 kabupaten/kota yaitu Gowa, Maros, Pangkep, Barru, Soppeng, Sidrap, Luwu, Luwu Timur, dan Pare Pare, dengan TPT pada level menengah yang mencerminkan dinamika ketenagakerjaan wilayah penyangga perkotaan dan kota madya berkembang. *Cluster 3* hanya beranggotakan Kota Makassar dan Kota Palopo yang memiliki TPT jauh di atas rata-rata provinsi secara konsisten, ditunjukkan oleh nilai jarak ke *Centroid 3* yang sangat kecil dibandingkan ke *centroid* lainnya. Meskipun keduanya merupakan pusat ekonomi di Sulawesi Selatan, tingginya TPT dapat dijelaskan melalui fenomena urbanisasi tanpa industrialisasi (*jobless urbanization*), dimana arus migrasi tenaga kerja dari kabupaten sekitar tidak diimbangi oleh pertumbuhan sektoral formal yang memadai [13].

Perbedaan karakteristik antar cluster ini memiliki implikasi penting bagi kebijakan ketenagakerjaan daerah. *Cluster 1* memerlukan kebijakan peningkatan produktivitas tenaga kerja di sektor pertanian dan perdesaan, *Cluster 2* membutuhkan penguatan sektor industri dan jasa untuk mengimbangi pertumbuhan angkatan kerja, sedangkan *Cluster 3* memerlukan intervensi yang lebih intensif berupa perluasan lapangan kerja formal, peningkatan keterampilan tenaga kerja, serta penguatan ekosistem kewirausahaan di wilayah perkotaan.

Untuk memperkuat hasil interpretasi hasil pengelompokan, dilakukan visualisasi skor PCA untuk seluruh kabupaten/kota disajikan pada Gambar 4 yang hasil *cluster*-nya diperoleh dari Tabel 9.



Gambar 4. Visualisasi PCA dengan Hasil *K-Means*

Berdasarkan visualisasi pada Gambar 4, sebagian besar titik data terkonsentrasi di sekitar pusat koordinat yang menandakan bahwa mayoritas kabupaten/kota memiliki karakteristik pengangguran yang tidak terlalu berbeda satu sama lain. Terdapat dua titik yang terpisah jauh ke arah kanan pada sumbu PC1, yaitu Kota Makassar dan Kota Palopo, yang secara visual mengkonfirmasi keunikan kedua wilayah ini. Dengan pewarnaan berdasarkan hasil *cluster*, terlihat jelas bahwa *K-Means* berhasil memisahkan tiga kelompok wilayah yang tidak saling tumpang tindih dalam ruang PCA 2D, sehingga memvalidasi kesesuaian antara hasil reduksi dimensi PCA dengan hasil pengelompokan *K-Means*.

Selanjutnya dilakukan perbandingan hasil antara *K-Means* tanpa PCA dengan *K-Means* berbasis PCA. Perbandingan dilakukan menggunakan dua metrik evaluasi, yaitu *Silhouette Score* dan *Davies-Boulding Index* (DBI). *Silhouette Score* mengukur kohesi dan separasi *cluster*, di mana nilai yang lebih mendekati 1 menunjukkan kualitas *clustering* yang lebih baik. Sebaliknya, *Davies-Boulding Index* mengukur rata-rata kemiripan antar *cluster*, di mana nilai yang lebih kecil mengindikasikan pemisahan *cluster* yang lebih optimal. Hasil perbandingan disajikan pada Tabel 10.=

Tabel 10. Perbandingan Hasil Cluster *K-Means* vs PCA + *K-Means*

Metode	<i>Silhouette Score</i>	<i>Davies-Boulding Index</i>
<i>K-Means</i>	0,4660	0,6440
PCA + <i>K-Means</i>	0,5031	0,5876

Berdasarkan Tabel 10, metode PCA + *K-Means* menghasilkan *Silhouette Score* lebih tinggi dibandingkan dengan *K-Means* tanpa PCA, yaitu 0,5031 berbanding 0,4660. Selain itu, nilai *Davies-Boulding Index* pada PCA + *K-Means* lebih rendah dibandingkan *K-Means* tanpa PCA, yaitu 0,5876 berbanding 0,6440. Kedua metrik ini secara konsisten menunjukkan bahwa integrasi PCA sebelum proses *clustering* menghasilkan kelompok yang lebih kohesif secara internal dan lebih terpisah antar kelompok. Hal ini mengonfirmasi bahwa reduksi dimensi dengan PCA efektif dalam mengeliminasi redundansi informasi akibat multikolinieritas yang tinggi antar variabel tahun pengamatan.

4. KESIMPULAN

Penelitian ini menerapkan analisis komponen utama (PCA) dan *K-Means clustering* pada data Tingkat Pengangguran Terbuka (TPT) 24 kabupaten/kota di Provinsi Sulawesi Selatan tahun 2019–2023. Hasil analisis PCA menghasilkan komponen utama yang mampu meringkas

informasi dari lima variabel tahun pengamatan dengan tetap mempertahankan sebagian besar varians data, dan skor komponen utama tersebut selanjutnya digunakan sebagai input dalam proses klasterisasi. Hasil *K-Means clustering* membagi 24 kabupaten/kota ke dalam tiga cluster, yaitu *Cluster 1* yang terdiri dari 13 kabupaten/kota dengan karakteristik TPT rendah dan stabil, *Cluster 2* yang terdiri dari 9 kabupaten/kota dengan TPT pada level menengah, dan *Cluster 3* yang terdiri dari Kota Makassar dan Kota Palopo dengan TPT tertinggi dan paling jauh menyimpang dari rata-rata provinsi. Temuan ini menunjukkan bahwa terdapat disparitas ketenagakerjaan yang signifikan antar wilayah di Sulawesi Selatan, sehingga diperlukan pendekatan kebijakan ketenagakerjaan yang berbeda dan disesuaikan dengan karakteristik masing-masing kelompok wilayah.

REFERENSI

- [1] Badan Pusat Statistik Provinsi Sulawesi Selatan. (2023). Agustus 2023: Tingkat Pengangguran Terbuka (TPT) Sulawesi Selatan Sebesar 4,33 Persen. BPS Sulawesi Selatan. <https://sulsel.bps.go.id/pressrelease/2023/11/06/759>
- [2] Selma L, Rahim A, Astuty S, Irwandi I, & Syafri M. Determinan Faktor-Faktor Yang Mempengaruhi Pengangguran Terbuka di Berbagai Kabupaten/Kota di Sulawesi Selatan. *Paradoks: Jurnal Ilmu Ekonomi*. 2026; 9(1):805–815.
- [3] Samsir A, Rahman A. Menelusuri Ketimpangan Distribusi Pendapatan Kabupaten dan Kota. *Jurnal EcceS : Economics, Sosial, dan Development Studies*. 2018; 5(1):22-42.
- [4] Rosyada IA, Utari DT. Penerapan Principal Component Analysis untuk reduksi variable pada algoritma K-Means Clustering. *Jambura J. Probab. Stat*. 2024; 5(1):6-13.
- [5] Dewi S, Pakereng MA. Implementasi Principal Component Analysis pada K-Means untuk Klasterisasi Tingkat Pendidikan Penduduk Kabupaten Semarang. *Jurnal Ilmiah Penelitian dan Pembelajaran Informatika*. 2023; 8(4): 1186-1195.
- [6] Akramunnisa, Fajriani. K-Means Clustering Analysis pada Persebaran Tingkat Pengangguran Kabupaten/Kota di Sulawesi Selatan. *Jurnal Varian*. 2020; 3(2): 103-112.
- [7] Pratama IWP. Standarisasi Z-Score sebagai Pendekatan Alternatif dalam Evaluasi Prestasi Akademik Mahasiswa: Studi Kasus di Politeknik eLBajo Commodus. *Jurnal Penelitian Terapan Mahasiswa*. 2023; 1(2): 77-85.
- [8] Purnama DI, Sihombing PR. Perbandingan Analisis Komponen Utama dan Robust PCA (ROBPCA). *Jurnal Bayesian : Jurnal Ilmiah Statistika dan Ekonometrika*. 2021; 1(1): 67-76.
- [9] Sa'diyah K, Primadhieta KD, Rosyid HA. Clustering Wilayah Pulau Jawa Berdasarkan Indikator Sosial Ekonomi Menggunakan Metode K-Means. *JATI : Jurnal Mahasiswa Teknik Informatika*. 2026; 10(1): 1032-1036.
- [10] Ariansyah B, Khaira U, Abidin Z. Penerapan K-Means Clustering untuk Mengelompokkan Data Industri Kecil Menengah di Provinsi Jambi. *Jurnal Teknologi Sistem Informasi*. 2025; 6(2): 359-371.
- [11] Suyal M, Sharma S. A Review on Analysis of K-Means Clustering Machine Learning Algorithm based on Unsupervised Learning. *Journal of Artificial Intelligence and Systems*. 2024; 6(1): 85-95.
- [12] Tobing RS, Sigalingging OH, Sinaga RK. Penerapan Principal Component Analysis (PCA) untuk Reduksi Dimensi dan Pemetaan Karakteristik Nutrisi pada Produk Makanan Kemasan di Indonesia. *Jurnal Matematika, Ilmu Pengetahuan Alam, Kebumihan dan Angkasa*. 2026; 4(1): 20-30.
- [13] Fatimah E, Anwar AC. Determination of Urban Sprawl Phenomenon in Pekalongan City And Its Surrounding. *Jurnal Pengembangan Kota*. 2023; 11(1): 49-57.