

Comparison of District/City *Clusters* in West Sumatra Province 2019–2025 Based on Labor Indicators Using *K-Means Method*

Naila Maretania, Mellisa Ayuningtyas, dan Zilrahmi*

Departemen Statistika, Universitas Negeri Padang, Padang, Indonesia

*Corresponding author: zilrahmi@fmipa.unp.ac.id

Submitted : 27 April 2026

Revised : 29 Mei 2026

Accepted : 30 Mei 2026

ABSTRACT

This study is motivated by the differences in labor conditions among regencies/cities in West Sumatra Province, as indicated by the Open Unemployment Rate (OUR) and the Labor Force Participation Rate (LFPR). In addition, the impact of the COVID-19 pandemic and the economic recovery process during the 2019–2025 period are assumed to have caused changes in labor characteristics across regions. However, the patterns of similarities and differences in labor conditions among regions have not been clearly identified, making it necessary to conduct a regional clustering analysis based on labor characteristics. This study aims to analyze the clustering of regencies/cities in West Sumatra Province based on the OUR and LFPR indicators during 2019–2025. The data used were obtained from the Central Statistics Agency, covering 19 regencies/cities. The analytical method applied was K-Means clustering using Euclidean distance, while cluster validation was conducted using the Silhouette Coefficient. This study used two clusters to facilitate the interpretation of results. The findings show that the regencies/cities in West Sumatra Province were divided into two clusters with different characteristics. Cluster 1 represents regions with better labor conditions, characterized by lower OUR and higher LFPR, while Cluster 2 represents regions with relatively poorer labor conditions, characterized by higher OUR and lower LFPR. Cluster membership changed from year to year, indicating dynamic labor conditions across regions. The results of this study are expected to serve as a basis for formulating more targeted labor policies according to the characteristics of each region.

Keywords *Clustering, Employment, K-Means, Regencies/cities, Silhouette Coefficient.*



This is an open access article under the Creative Commons 4.0 Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. ©2022 by author and Universitas Negeri Padang.

I. PENDAHULUAN

Ketenagakerjaan mencakup seluruh aspek yang berkaitan dengan tenaga kerja sebelum, selama, dan setelah masa kerja (UU No. 13 Tahun 2003 tentang Ketenagakerjaan). Kondisi ini mengindikasikan bahwa ketenagakerjaan mencakup aktivitas bekerja sekaligus meliputi proses persiapan tenaga kerja. Kondisi ketenagakerjaan yang baik akan tercermin dari tersedianya lapangan kerja yang memadai serta kemampuan tenaga kerja untuk berpartisipasi secara aktif dalam kegiatan ekonomi. Dalam pembangunan, peningkatan kualitas Sumber Daya Manusia (SDM) merupakan salah satu faktor utama selain pertumbuhan ekonomi (Adriani et al., 2019). SDM yang berkualitas mampu meningkatkan produktivitas, memperluas kesempatan kerja, dan memperkuat daya saing daerah. Kualitas SDM memiliki keterkaitan yang erat dengan kondisi ketenagakerjaan di suatu daerah. Oleh karena itu, ketenagakerjaan menjadi indikator penting dalam menilai keberhasilan pembangunan.

Menurut Badan Pusat Statistik (BPS, 2025a), indikator utama ketenagakerjaan adalah Tingkat Pengangguran Terbuka (TPT) dan Tingkat Partisipasi Angkatan Kerja (TPAK). TPT menggambarkan persentase penduduk yang menganggur dalam angkatan kerja, sedangkan TPAK menggambarkan tingkat partisipasi penduduk usia kerja (BPS, 2025b). Tingginya TPT mencerminkan rendahnya penyerapan tenaga kerja. Pada tahun 2020, TPT Sumatera Barat sebesar 6,88 persen dan menempatkan Provinsi Sumatera Barat ke dalam 10 provinsi dengan TPT tertinggi di Indonesia (BPS, 2020a). Sementara itu, rendahnya TPAK menunjukkan bahwa sebagian penduduk usia kerja belum berpartisipasi secara aktif dalam kegiatan ekonomi, sehingga dapat mengindikasikan kurang optimalnya pemanfaatan potensi tenaga kerja di suatu daerah. Pada tahun 2021, TPAK Sumatera Barat tercatat sebesar 67,72 persen, yang merupakan nilai

terendah selama periode 2019–2025, sehingga menunjukkan bahwa partisipasi angkatan kerja pada tahun tersebut berada pada kondisi yang paling rendah dibandingkan tahun-tahun lainnya (BPS, 2020b).

Kabupaten/kota di Provinsi Sumatera Barat mempunyai karakteristik wilayah yang beragam, baik dari sisi pertumbuhan ekonomi, tingkat urbanisasi, maupun perkembangan sektor-sektor unggulan antar kabupaten/kota. Dampak pandemi COVID-19 dan pemulihan ekonomi selama 2019-2025 turut mempengaruhi kondisi ketenagakerjaan di Provinsi Sumatera Barat. Perbedaan karakteristik dan dinamika ini menunjukkan bahwa setiap daerah memiliki karakteristik dan permasalahan ketenagakerjaan yang berbeda. Namun, hingga saat ini belum terdapat pengelompokan kabupaten/kota di Provinsi Sumatera Barat berdasarkan karakteristik ketenagakerjaan yang mempertimbangkan indikator ketenagakerjaan yaitu TPT dan TPAK secara bersamaan dalam periode 2019-2025. Akibatnya kondisi ketenagakerjaan antar daerah sulit dibandingkan secara menyeluruh, sehingga pemerintah daerah berpotensi mengalami kesulitan dalam menentukan prioritas kebijakan ketenagakerjaan yang sesuai dengan karakteristik masing-masing wilayah. Perubahan kondisi ketenagakerjaan akibat pandemi COVID-19 dan proses pemulihan ekonomi diduga menyebabkan terjadinya pergeseran karakteristik ketenagakerjaan antar kabupaten/kota dari tahun ke tahun. Namun, pola kesamaan maupun perubahan jkarakteristik tersebut belum terindikasi secara jelas. Oleh karena itu, diperlukan suatu metode analisis yang mampu mengidentifikasi pola kesamaan antar wilayah secara komprehensif. Hasil identifikasi tersebut dapat menjadi dasar dalam pengelompokkan wilayah yang memiliki karakteristik yang serupa, sehingga dapat mendukung penyusunan kebijakan ketenagakerjaan yang lebih tepat sesuai dengan kondisi tiap daerah (Maliqi et al., 2022).

Penelitian ini bertujuan untuk mengelompokkan kabupaten/kota di Provinsi Sumatera Barat berdasarkan indikator ketenagakerjaan menggunakan analisis *cluster*. Metode yang digunakan adalah analisis *cluster K-Means*. *K-Means* merupakan salah satu metode *data mining* yang menerapkan proses pemodelan tanpa supervise (*unsupervised learning*) dan digunakan untuk mengelompokkan data dengan sistem partisi. Metode ini bertujuan untuk meminimalkan variasi data yang ada di dalam satu *cluster* dan memaksimalkan variasi antar *cluster* (Agusta, 2007). Dalam penerapannya, Metode *K-Means* mengelompokkan data yang memiliki tingkat kemiripan ke dalam *cluster* yang sama. Tingkat kemiripan tersebut diukur berdasarkan jarak antar objek, sehingga objek dengan jarak paling dekat akan ditempatkan ke *cluster*. Hasil pengelompokkan ini memberikan informasi mengenai perbedaan dan persamaan kondisi ketenagakerjaan antar daerah, sehingga dapat digunakan sebagai dasar dalam memahami karakteristik tiap wilayah serta penyusunan kebijakan yang lebih tepat sesuai dengan kondisi daerah masing masing. Penelitian ini juga akan disajikan dalam bentuk peta visualisasi setiap tahun untuk menggambarkan persebaran *cluster* antar kabupaten/kota secara spasial, sehingga perubahan dan pola pengelompokkan wilayah dapat terlihat jelas.

Penelitian terdahulu oleh Mirkin (2011) menjelaskan bahwa metode *K-Means* efektif digunakan untuk mengelompokkan data berdasarkan ukuran jarak antar objek, khususnya pada data memiliki pola distribusi yang jelas serta kelompok yang relatif terpisah dengan baik. Dalam kondisi tersebut, algoritma *K-Means* dapat mengenali pola data dan membentuk *cluster* secara optimal. Hal ini selaras dengan tujuan penelitian ini, yaitu untuk mengelompokkan Provinsi di Indonesia berdasarkan karakteristik indikator ketenagakerjaan seperti TPT dan TPAK. Dengan adanya perbedaan nilai TPT dan TPAK antar provinsi, metode *K-Means* diharapkan mampu mengelompokkan provinsi-provinsi tersebut ke dalam beberapa *cluster* yang memiliki karakteristik kondisi ketenagakerjaan yang serupa.

Penelitian yang dilakukan oleh Jain (2010) menjelaskan bahwa metode *k-means* memiliki tingkat kompleksitas komputasi yang relatif lebih rendah dibandingkan dengan metode clusterisasi yang lain yang lebih kompleks, seperti Gaussian Mixture Model (GMM). Hal ini menunjukkan bahwa *k-means* lebih cepat dan sederhana dalam penerapannya, khusus ketika digunakan pada dataset berukuran besar. Keunggulan tersebut menjadi salah satu pertimbangan utama pada dalam penggunaan *k-means* pada penelitian ini. Dengan jumlah provinsi serta variasi data TPT dan TPAK yang cukup beragam, metode *k-means* mampu mengelompokkan data secara efisien tanpa memerlukan proses komputasi yang kompleks, sehingga mempermudah analisis pola ketenagakerjaan di Indonesia.

II. METODE PENELITIAN

Penelitian ini merupakan penelitian terapan yang memanfaatkan data sekunder yang diperoleh dari website resmi Badan Pusat Statistik Provinsi Sumatera Barat (<https://sumbar.bps.go.id>) dengan cakupan data dari 19 kabupaten/kota di Provinsi Sumatera Barat. Dalam penelitian ini terdapat dua variabel yang merupakan indikator utama ketenagakerjaan. Adapun 2 variabel tersebut yaitu Tingkat Pengangguran terbuka (X_1) dan Tingkat Partisipasi Angkatan Kerja (X_2).

Menurut (Aggarwal, 2015), tahapan analisis yang diterapkan adalah sebagai berikut:

1. Menentukan jumlah *cluster*. Pada penelitian ini, dipilih dua *cluster* ($k = 2$) untuk mempermudah interpretasi dan perbandingan karakteristik data antar tahun. Selain itu, pemilihan ini juga mengacu pada penelitian sebelumnya yang menunjukkan bahwa dua *cluster* mampu memberikan gambaran yang sederhana namun tetap jelas (Ramadhan et al., 2024).
2. Menentukan *Centroid* awal secara acak. *Centroid* awal dibangkitkan secara acak dari data kabupaten/kota, sehingga setiap objek memiliki peluang yang sama. Pemilihan acak ini bertujuan menghindari bias dan memungkinkan proses iterasi *K-Means* berlangsung secara objektif hingga konvergen.
3. Kelompokkan objek ke *centroid* terdekat berdasarkan jarak terkecil objek ke *centroid*. Menurut (Aggarwal, 2015), jarak dapat dihitung dengan menggunakan jarak *Euclidean* menggunakan rumus:

$$d_{ik} = \sqrt{\sum_{i=1}^p (x_{ik} - x_{kj})^2} \quad i = 1,2, \dots, n \text{ dan } k = 1,2, \dots, m \quad (1)$$

Pada persamaan tersebut, d_{ik} menyatakan jarak *Euclidean* objek ke- i dan objek ke- k , x_{ik} merupakan nilai objek data ke- i pada peubah ke- k , x_{kj} adalah nilai *centroid* ke- k pada variabel ke- j , p menunjukkan banyaknya variabel yang diamati, n adalah banyaknya objek yang diamati, dan m merupakan banyaknya *centroid* yang terbentuk.

4. Setiap objek data dikelompokkan ke dalam *cluster* dengan jarak paling dekat terhadap *centroid*.
5. Melakukan pembaruan nilai *centroid*. *Centroid* terbaru dihitung berdasarkan rata-rata dalam masing-masing *cluster* dengan rumus:

$$C = \frac{\sum x_i}{N} \quad i = 1,2, \dots, n \quad (2)$$

Dimana C adalah nilai rata-rata *cluster*, $\sum x_i$ adalah jumlah seluruh nilai data yang termasuk dalam *cluster* tersebut, N adalah jumlah seluruh data dalam *cluster* tersebut dan i adalah banyak variabel yang diamati.

6. Mengulangi proses 2 sampai 5 hingga hasil konvergen.
7. Setelah proses pengelompokan selesai untuk satu tahun pengamatan, ulangi langkah yang sama diterapkan pada setiap tahun penelitian yaitu 2019-2025.
8. Selanjutnya hasil dirangkum pengelompokan setiap tahun disajikan dalam bentuk tabel untuk mempermudah identifikasi perpindahan kelompok kabupaten/kota antar tahun..
9. Selanjutnya dilakukan validasi hasil analisis *cluster* menggunakan metode *Silhouette Coefficient*. *Silhouette Coefficient* merupakan metode evaluasi yang bertujuan untuk menilai kualitas *cluster* dengan mengombinasikan dua pendekatan, yaitu kohesi (kesatuan dalam *cluster*) dan separasi (pemisahan antar *cluster*) (Utami et al., 2023). Pendekatan ini termasuk salah satu cara yang paling sering diterapkan untuk memvalidasi hasil *cluster*. Adapun rumus menghitung *silhouette Coefficient* (Utami et al., 2023):

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (3)$$

Dimana

$s(i)$ adalah skor *silhouette Coefficient* untuk objek ke- i

$a(i)$ adalah rata-rata antar objek ke- i dengan seluruh objek yang berada dalam *cluster* yang sama

$b(i)$ adalah rata-rata jarak antar objek ke- i dengan objek pada *cluster* terdekat lainnya.

Nilai *Silhouette Coefficient* memiliki nilai antara -1 hingga 1, apabila nilainya semakin mendekati 1 mengindikasikan kualitas *clustering* yang semakin baik. Menurut Rousseeuw (1987) kriteria dalam pengukuran *Silhouette Coefficient* disajikan pada Tabel 1.

Tabel 1. Ukuran *Silhouette Coefficient*

<i>Silhouette Coefficient</i>	Kriteria <i>Cluster</i>
$0,7 < SC \leq 1,0$	Struktur <i>cluster</i> sangat baik
$0,5 < SC \leq 0,7$	Struktur <i>cluster</i> baik
$0,25 < SC \leq 0,5$	Struktur <i>cluster</i> lemah
$SC \leq 0,25$	Struktur <i>cluster</i> buruk

10. Membuat peta untuk masing-masing tahun untuk memvisualisasikan sebaran *cluster* secara spasial dan memperjelas dinamika perubahan pengelompokan kabupaten/kota di Provinsi Sumatera Barat.

III. HASIL DAN PEMBAHASAN

A. Analisis Cluster K-Means

Dalam analisis *cluster*, jumlah *cluster* dapat ditentukan berdasarkan pertimbangan teoritis, karakteristik data, maupun tujuan penelitian. Pada penelitian ini jumlah *cluster* ditetapkan sebanyak dua ($k=2$). Pemilihan dua *cluster* dilakukan dengan pertimbangan untuk mempermudah interpretasi hasil serta memudahkan perbandingan karakteristik data antar tahun pengamatan. Dengan dua kelompok, perbedaan pola antar tahun dapat dianalisis secara lebih jelas dan sistematis. Adapun 2 variabel yang digunakan yaitu Tingkat Pengangguran terbuka (X_1) dan Tingkat Partisipasi Angkatan Kerja (X_2).

Langkah selanjutnya setelah menentukan jumlah *cluster* didapatkan adalah pilih 2 titik data secara acak yang akan digunakan sebagai *centroid*.

Tabel 2. Centroid Awal

Cluster	Variabel	TAHUN						
		2019	2020	2021	2022	2023	2024	2025
1	x_1	6,84	8,2	6,38	5	3,72	5,55	4,64
	x_2	69,83	70,57	68,05	70,85	70,12	73,54	74,02
2	x_1	7,06	8,35	5,15	3,9	4,98	3,62	3,87
	x_2	62,51	66,77	66,51	69,46	70,12	71,16	70,32

Berdasarkan Tabel 2, diperoleh nilai *centroid* awal dari masing-masing *cluster*. Kemudian hitung jarak antara setiap objek non-*centroid* dan *centroid* menggunakan ukuran jarak Euclidean. Proses analisis *cluster K-Means* dilakukan secara iteratif hingga posisi *centroid* mencapai kondisi tetap atau tidak mengalami perubahan. Hasil *Centroid* akhir dapat dilihat di Tabel 3.

Tabel 3. Centroid Akhir

Cluster	Variabel	TAHUN						
		2019	2020	2021	2022	2023	2024	2025
1	x_1	4,55	4,69	4,22	4,81	4,14	3,98	3,84
	x_2	71,14	73,54	73,15	73,05	74,97	76,52	76,5
2	x_1	6,56	7,44	6,33	6,06	5,68	5,48	5,49
	x_2	63,8	67,54	66,27	65,43	68,18	69,08	70,24

Berdasarkan Tabel 3, diperoleh nilai *centroid* akhir dari proses analisis *cluster K-Means* untuk tahun 2019-2025 yang mencerminkan karakteristik masing-masing *cluster*. *Cluster* 1 memiliki kondisi ketenagakerjaan yang lebih baik dibandingkan *cluster* 2, karena ditandai dengan TPT yang lebih rendah dan TPAK yang lebih tinggi dibandingkan *cluster* 2. Sementara itu, *cluster* 2 menunjukkan kondisi ketenagakerjaan yang relatif baik dengan TPT yang lebih tinggi dan TPAK yang lebih rendah dibandingkan *cluster* 1. Hasil pengelompokan Kabupaten/kota di Provinsi Sumatera Barat yang mengalami perpindahan *cluster* berdasarkan indikator ketenagakerjaan menggunakan analisis *K-Means* disajikan pada Tabel 4.

Tabel 4. Hasil cluster Tahun 2019-2025

Kabupaten/Kota	Tahun						
	2019	2020	2021	2022	2023	2024	2025
Kab. Pesisir Selatan	1	2	2	2	2	2	2
Kab. Sijunjung	1	1	1	2	2	2	1
Kab. Tanah Datar	1	1	2	1	2	2	1
Kab. Padang Pariaman	2	2	2	1	2	2	1
Kab. Agam	1	1	2	1	2	2	2

Kab. Pasaman	1	1	2	1	1	2	1
Kota Solok	2	2	2	1	2	2	2
Kota Sawahlunto	1	1	2	1	2	1	1
Kota Padang Panjang	1	1	2	2	2	2	2
Kota Bukittinggi	2	1	2	2	2	2	2
Kota Payakumbuh	1	2	1	1	1	2	2
Kota Pariaman	1	2	2	2	2	2	1

Tabel 4 menunjukkan dinamika perpindahan *cluster* kabupaten/kota di Provinsi Sumatera Barat pada periode 2019–2025 dengan karakteristik *cluster* yang berbeda, yaitu *Cluster 1* memiliki kondisi ketenagakerjaan yang lebih baik dibandingkan *cluster 2*, karena ditandai dengan TPT yang lebih rendah dan TPAK yang lebih tinggi dibandingkan *cluster 2*. Sementara itu, *cluster 2* menunjukkan kondisi ketenagakerjaan yang relatif baik dengan TPT yang lebih tinggi dan TPAK yang lebih rendah dibandingkan *cluster 1*. Meskipun nilai *centroid* pada setiap tahun dapat mengalami perubahan, karakteristik umum dari masing masing *cluster* tetap menunjukkan pola yang serupa selama periode penelitian. Secara umum, terlihat bahwa sebagian besar kabupaten/kota mengalami perpindahan dari *Cluster 1* ke *Cluster 2*, yang mengindikasikan adanya penurunan kondisi ketenagakerjaan, ditandai dengan meningkatnya tingkat pengangguran dan menurunnya partisipasi angkatan kerja.

Beberapa daerah seperti Kab. Pesisir Selatan, Kab. Agam, Kota Padang Panjang, Kota Solok, dan Kota Bukittinggi menunjukkan pergeseran menuju *Cluster 2* dan cenderung bertahan pada *cluster* tersebut hingga tahun-tahun akhir, yang mencerminkan kondisi ketenagakerjaan yang kurang menguntungkan. Sementara itu, daerah seperti Kab. Sijunjung, Kab. Tanah Datar, Kab. Padang Pariaman, Kab. Pasaman Kota Sawahlunto, dan Kota Pariaman mengalami perpindahan *cluster* yang lebih fluktuatif, yang menunjukkan kondisi ketenagakerjaan yang tidak stabil dari waktu ke waktu dan terdapat daerah yang sempat berada pada kondisi yang lebih baik (*Cluster 1*) namun kemudian berpindah ke *Cluster 2*, seperti Kota Payakumbuh. Hal ini menunjukkan adanya perubahan kondisi ekonomi yang berdampak pada penyerapan tenaga kerja. Secara keseluruhan, tabel ini mengindikasikan bahwa mayoritas kabupaten/kota cenderung mengalami pergeseran ke arah kondisi ketenagakerjaan yang kurang baik selama periode pengamatan.

Di sisi lain, terdapat daerah yang menunjukkan konsistensi dalam keanggotaan *cluster*. Kabupaten Kepulauan Mentawai, Kabupaten Solok Selatan, Kabupaten Solok, Kabupaten Lima Puluh Kota, dan Kabupaten Dharmasraya tetap berada pada *Cluster 1* selama periode pengamatan, yang mencerminkan kondisi ketenagakerjaan yang relatif lebih baik dan stabil. Di sisi lain, Kabupaten Pasaman Barat, dan Kota Padang, tetap berada pada *Cluster 2*, yang mengindikasikan bahwa daerah-daerah tersebut secara konsisten menghadapi tantangan ketenagakerjaan berupa tingkat pengangguran yang relatif tinggi dan partisipasi angkatan kerja yang lebih rendah.

B. Validasi

Proses validasi merupakan tahap akhir dalam penelitian yang bertujuan untuk menilai kualitas kinerja model *clustering*. Validasi dilakukan dengan menghitung tingkat kedekatan antar data, baik jarak antar data point terhadap *centroid cluster* maupun jarak antar *cluster*. Dalam penelitian ini, digunakan metode *Silhouette Coefficient* sebagai ukuran untuk menilai kualitas model yang dihasilkan. *Silhouette Coefficient* digunakan untuk seberapa baik suatu data tergabung dalam *cluster*-nya dibandingkan terhadap *cluster* yang lain. Nilai yang semakin mendekati nilai 1, mengindikasikan bahwa kualitas model *clustering* semakin baik. Berdasarkan hasil perhitungan, nilai *Silhouette Coefficient* disajikan pada Tabel 5.

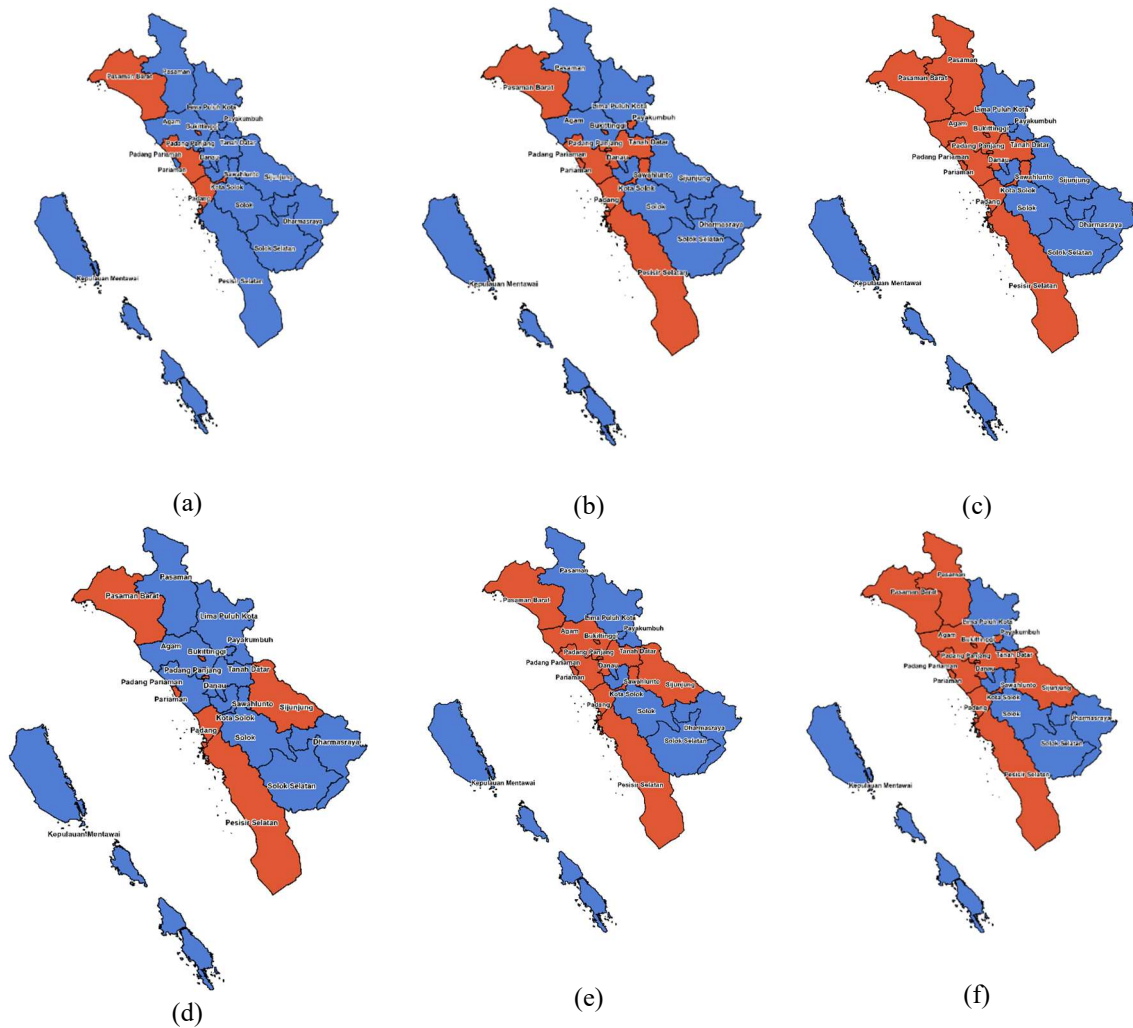
Tabel 5. Hasil nilai *Silhouette Coefficient*

Tahun	<i>Silhouette Coefficient</i>
2019	0,54
2020	0,39
2021	0,47
2022	0,46
2023	0,45
2024	0,51
2025	0,45

Berdasarkan Tabel 5, nilai *Silhouette Coefficient* pada periode 2019–2025 berada pada kisaran 0,39 hingga 0,54 yang menunjukkan bahwa kualitas hasil *clustering* tergolong cukup baik (moderate), karena sebagian besar nilainya mendekati 0,5 sehingga mengindikasikan bahwa objek dalam *cluster* sudah cukup homogen dan memiliki pemisahan yang cukup jelas dengan *cluster* lain.

C. Visualisasi

Peta visualisasi hasil pengelompokan disajikan pada Gambar 1, yang merupakan hasil visualisasi menggunakan *software RStudio* berdasarkan data hasil pemodelan yang telah dilakukan. Visualisasi tersebut menampilkan persebaran *cluster*, di mana warna biru menunjukkan daerah yang termasuk *cluster* 1 dengan TPT rendah dan TPAK tinggi, sedangkan warna merah menunjukkan *cluster* 2 dengan TPT tinggi dan TPAK rendah.



- BPS, Provinsi Sumatera Barat. (2020a). *Tingkat Pengangguran Terbuka (TPT) Menurut Kabupaten/Kota—Tabel Statistik*. <https://sumbar.bps.go.id/id/statistics-table/2/MTM5IzI=/tingkat-pengangguran-terbuka--tpt--menurut-kabupaten-kota--persen-.html>
- BPS, Provinsi Sumatera Barat. (2025b). *Keadaan Angkatan Kerja Provinsi Sumatera Barat Agustus 2025*. <https://sumbar.bps.go.id/id/publication/2026/03/31/6a391c9245eccc1a88b31354/keadaan-angkatan-kerja-provinsi-sumatera-barat-agustus-2025.html>
- Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters, Award winning papers from the 19th International Conference on Pattern Recognition (ICPR)*, 31(8), 651–666. <https://doi.org/10.1016/j.patrec.2009.09.011>
- Maliqi, R., Falgenti, K., Priani, S., Fajrul, F., Muhammad, S., & Dwi Satria, N. (2022). *Perbandingan Tingkat Pengangguran Terbuka Provinsi di Indonesia Berbasis Metode K-Means Clustering*. 2. <https://paperity.org/p/320737135/perbandingan-tingkat-pengangguran-terbuka-provinsi-di-indonesia-berbasis-metode-k-means>
- Mirkin, B. (2011). Choosing the number of clusters. *WIREs Data Mining and Knowledge Discovery*, 1(3), 252–260. <https://doi.org/10.1002/widm.15>
- Ramadhan, A. I., Atika, P. D., & Ramdhania, K. F. (2024). Analisis Clustering K-Means untuk Pemetaan Tingkat Pengangguran Terbuka di Provinsi-Provinsi Indonesia Tahun 2013-2023. *Journal of Students' Research in Computer Science*, 5(2), 109–122. <https://doi.org/10.31599/wbpydb62>
- Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, 53–65. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7)
- Utami, I. T., Suryaningrum, F., & Ispriyanti, D. (2023). K-Means Cluster Count Optimization With Silhouette Index Validation And Davies Bouldin Index (Case Study: Coverage Of Pregnant Women, Childbirth, And Postpartum Health Services In Indonesia In 2020). *BAREKENG: Jurnal Ilmu Matematika Dan Terapan*, 17(2), 0707–0716. <https://doi.org/10.30598/barekengvol17iss2pp0707-0716>
- UU No. 13 Tahun 2003 tentang Ketenagakerjaan. (t.t.). Diambil 25 April 2026, dari <https://peraturan.bpk.go.id/Details/43013>