

ABSTRACT

The Open Unemployment Rate (OUR) serves as a critical indicator for assessing labor market conditions and regional economic stability. The characteristics of OUR across provinces in Indonesia are dynamic, as they change over time, necessitating analytical methods capable of capturing similarities in time-series movement patterns rather than merely comparing values at a single point in time. This study clusters provinces in Indonesia based on 22 time-series data points from 2015 to 2025. The research implements two clustering algorithms, K-Medoids and Fuzzy K-Medoids, using two distance metrics, Euclidean and Dynamic Time Warping (DTW), with cluster validation performed using the Silhouette Coefficient. The selection of both clustering methods is based on their non-parametric nature, ensuring the grouping process is not constrained by specific data distribution assumptions. The OUR data was analyzed as time-series data, followed by clustering with variations in the number of clusters $k = 2, 3, 4$, and 5. The Silhouette Coefficient was used as an evaluation indicator to assess the compactness and separation of the resulting clusters. The results show that Fuzzy K-Medoids using the Dynamic Time Warping distance at $k = 2$ yielded the highest Silhouette value of 0.627. This indicates that Fuzzy K-Medoids is more effective at capturing similar patterns of change in unemployment rates across provinces.

Keywords: *Open Unemployment Rate, Time Series, K-Medoids, Fuzzy K-Medoids, Dynamic Time Warping, Euclidean, Silhouette Coefficient.*

DAFTAR ISI

HALAMAN JUDUL.....	ii
HALAMAN PENGESAHAN I.....	iii
HALAMAN PENGESAHAN II	iv
KATA PENGANTAR.....	v
ABSTRAK	vi
ABSTRACT	vii
DAFTAR ISI	viii
DAFTAR TABEL	xi
DAFTAR GAMBAR	xii
DAFTAR LAMPIRAN.....	xiii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	5
1.3 Batasan Masalah.....	6
1.4 Tujuan Penelitian.....	7
BAB II TINJAUAN PUSTAKA	8
2.1 Tingkat Pengangguran Terbuka.....	8
2.2 Clustering Data Deret Waktu	9
2.3 Konsep Jarak	14

2.3.1	<i>Euclidean</i>	15
2.3.2	<i>Dynamic Time Warping</i>	17
2.4	<i>K-Medoids</i>	22
2.5	<i>Fuzzy K-Medoids</i>	30
2.6	<i>Silhouette Index</i>	35
BAB III METODOLOGI PENELITIAN.....		40
3.1	Jenis dan Sumber Data	40
3.2	Variabel Penelitian.....	40
3.3	Tahapan Analisis Data	41
3.4	Diagram Alir Analisis Data	45
BAB IV HASIL DAN PEMBAHASAN.....		46
4.1	Deskripsi Data.....	46
4.2	Perbandingan Metode Jarak	48
4.2.1	<i>Euclidean</i>	48
4.2.2	<i>Dynamic Time Warping</i>	51
4.3	Klastering <i>K-Medoids</i>	56
4.3.1	Klastering <i>K-Medoids</i> dengan Jarak <i>Euclidean</i>	56
4.3.2	Klastering <i>K-Medoids</i> dengan Jarak <i>Dynamic Time Warping</i>	72
4.4	Klastering <i>Fuzzy K-Medoids</i>	73
4.4.1	Klastering <i>Fuzzy K-Medoids</i> dengan Jarak <i>Euclidean</i>	73
4.4.2	Klastering <i>Fuzzy K-Medoids</i> dengan Jarak <i>DTW</i>	82

4.5	Evaluasi Perbandingan	84
4.6	Interpretasi Klaster	85
BAB V PENUTUP		90
5.1	Kesimpulan	90
DAFTAR PUSTAKA		92
LAMPIRAN		95

DAFTAR TABEL

Tabel 1. Layout Data Penelitian.....	9
Tabel 2. Kriteria <i>Silhouette Coefficient</i>	39
Tabel 3. Data Penelitian	41
Tabel 4. Deskripsi Data.....	46
Tabel 5. Karakteristik Antarprovinsi Jarak <i>Euclidean</i>	50
Tabel 6. Karakteristik Antarprovinsi Jarak <i>Dynamic Time Warping</i>	54
Tabel 7. Perbandingan Jarak <i>Euclidean</i> dan <i>Dynamic Time Warping</i>	55
Tabel 8. Medoid Awal Jarak <i>Euclidean K-Medoids</i>	63
Tabel 9. Medoid Akhir Jarak <i>Euclidean K-Medoids</i>	70
Tabel 10. Medoid Awal Jarak DTW <i>K-Medoids</i>	72
Tabel 11. Medoid Akhir Jarak DTW <i>K-Medoids</i>	72
Tabel 12. Medoid Awal Jarak <i>Euclidean Fuzzy K-Medoids</i>	74
Tabel 13. Medoid Akhir Jarak <i>Euclidean Fuzzy K-Medoids</i>	82
Tabel 14. Medoid Awal Jarak <i>Dynamic Time Warping Fuzzy K-Medoids</i>	83
Tabel 15. Medoid Akhir Jarak <i>Dynamic Time Warping Fuzzy K-Medoids</i>	83
Tabel 16. Evaluasi Perbandingan Metode.....	84
Tabel 17. Provinsi Hasil Klasterisasi	86

DAFTAR GAMBAR

Gambar 1. Diagram Alir Penelitian	45
Gambar 2. Visualisasi Data TPT Setiap Provinsi Tahun 2015-2025.....	47
Gambar 3. Visualisasi Hasil Klaster	87

DAFTAR LAMPIRAN

Lampiran 1. Data Tingkat Pengangguran Terbuka	95
Lampiran 2. Statistika Deskriptif	96
Lampiran 3. Matriks Jarak <i>Euclidean</i>	98
Lampiran 4. Matriks Jarak <i>Dynamic Time Warping</i>	102
Lampiran 5. Provinsi Terdekat dan Terjauh Jarak <i>Euclidean</i>	106
Lampiran 6. Provinsi Terdekat dan Terjauh Jarak <i>Dynamic Time Warping</i>	107
Lampiran 7. <i>Syntax</i>	108

BAB I

PENDAHULUAN

1.1 Latar Belakang

Kemajuan dan kesejahteraan sebuah negara diukur dari bagaimana pertumbuhan dan kualitas pembangunan ekonomi di dalamnya. Visi pembangunan kini berfokus pada strategi melalui misi Tim Prabowo–Gibran (2024) Asta Cita guna mewujudkan target Indonesia emas 2045. Salah satu pilar dalam visi tersebut adalah menekankan penguatan pembangunan Sumber Daya Manusia (SDM), sains, teknologi, serta pendidikan untuk meningkatkan daya saing masyarakat. Urgensi penguatan daya saing ini menjadi relevan mengingat Indonesia masih menghadapi tantangan pada aspek ketenagakerjaan. Berdasarkan laporan *World Economic Outlook* dari *International Monetary Fund* (2024) pada April 2024 tingkat pengangguran di Indonesia tertinggi di ASEAN, sebesar 5,2% dan menurut Santika (2025) tetap bertahan di level tinggi 5% pada tahun 2025. Angka ini menempatkan Indonesia sebagai salah satu negara dengan tingkat pengangguran tertinggi di ASEAN.

Tingkat Pengangguran Terbuka (TPT) merupakan salah satu indikator krusial dalam mengukur kesehatan pasar tenaga kerja dan keberhasilan penciptaan lapangan kerja yang dirilis secara berkala oleh Badan Pusat Statistik (BPS). (Booklet Sakernas, 2025). Berdasarkan Survei Angkatan Kerja Nasional (Sakernas) yang dilakukan oleh Badan Pusat Statistik (BPS), pada Februari 2024 tercatat bahwa jumlah penganggur di Indonesia terus bertambah dari 7,20 juta orang menjadi 7,28 juta orang pada Februari 2025. Kondisi ini menunjukkan jumlah penganggur masih bertambah dan menjadi isu serius dalam ketenagakerjaan di

Indonesia dalam satu tahun terakhir. Provinsi – provinsi di Indonesia masih menunjukkan kesenjangan angka pengangguran. Pada tahun 2024 berdasarkan laporan dari Sakernas (2024) TPT tertinggi pada provinsi Banten, Riau, dan Jawa Barat, sementara provinsi Bali pada posisi TPT terendah, hal ini juga didukung oleh hasil penelitian klastering TPT pada tahun 2020 hingga 2024 yang dilakukan oleh Karim *et al.* (2024) bahwa Banten dan Jawa Barat masuk sebagai klaster dengan TPT tinggi. Pada tahun 2025 provinsi dengan TPT tertinggi dan terendah sedikit bergeser. Provinsi dengan TPT tertinggi ditempati oleh provinsi Papua, Kepulauan Riau, dan Jawa Barat, sementara provinsi dengan TPT terendah yaitu provinsi Bali dan Sulawesi Tengah. Kondisi ini menunjukkan bahwa permasalahan TPT di Indonesia tidak sekadar angka, tetapi juga bagaimana beban pengangguran tersebar tidak merata antarprovinsi. Melihat realitas tersebut, maka permasalahan pengangguran ini perlu dikaji lebih dalam, karena setiap provinsi memiliki karakteristik ekonomi, struktur sektor unggulan, kapasitas penciptaan lapangan pekerjaan, serta kualitas dan kuantitas SDM yang berbeda.

Penelitian pengelompokan data deret waktu telah dilakukan dengan berbagai metode dan metrik jarak. Penelitian terdahulu yang dilakukan oleh Aqsari *et al.* (2022) menyimpulkan bahwa *Dynamic Time Warping* (DTW) lebih efektif dalam mengelompokkan harga saham karena mampu menangkap kemiripan pola fluktuasi data deret waktu. Soleha *et al.* (2022) melakukan pengelompokan deret waktu juga terhadap 34 provinsi di Indonesia berdasarkan nilai ekspor nonmigas pada tahun 2016 hingga 2020 dengan membandingkan dua ukuran jarak yaitu *Euclidean* dan DTW. Hasil penelitian Soleha (2022) menunjukkan bahwa DTW menghasilkan klaster lebih baik dilihat dari nilai koefisien *silhouette* dan koefisien *cophenetic*

dibandingkan jarak *Euclidean* untuk ekspor nonmigas. Penggunaan DTW menurut Izakian *et al.* (2015) dinilai lebih unggul karena mampu menangani karakteristik data *time series* yang memiliki pergeseran waktu dan pola tidak sejajar.

Ding *et al.* (2008) menegaskan bahwa jarak *Euclidean* merupakan ukuran jarak yang sangat umum dipakai dalam analisis kluster karena mudah diimplementasikan. Penelitian Holder *et al.* (2023) menunjukkan bahwa *Dynamic Time Warping* (DTW) tidak selalu lebih unggul dibandingkan jarak *Euclidean*. Dalam eksperimen mereka, DTW memiliki kinerja lebih buruk dibandingkan jarak *Euclidean* pada klusterisasi deret waktu. Sejalan dengan penelitian tersebut, penelitian Izakian *et al.* (2015) memberikan hasil bahwa *clustering* berbasis *means* dan jarak *Euclidean* memiliki kinerja yang buruk dan lemah untuk sebagian besar kumpulan data, tetapi memberikan hasil yang baik untuk kumpulan data minyak zaitun dimana data tersebut pergeseran waktu dalam deret tidak terlalu terlihat. Pemilihan jarak pada data deret waktu tidak dapat langsung mengasumsikan bahwa *Euclidean* selalu lebih buruk dibanding *Dynamic Time Warping*, tetapi perlu diuji sesuai karakteristik data.

Izakian *et al.* (2015) menekankan bahwa pada pengelompokkan *fuzzy* data deret waktu, penggunaan pusat kluster berbasis medoid (*Fuzzy C-Medoids*) lebih sesuai ketika dengan *Dynamic Time Warping* karena tidak memerlukan perhitungan rata-rata. Berangkat dari temuan – temuan tersebut, penelitian ini mengevaluasi empat metode, yaitu *K-Medoids - Euclidean*, *K-Medoids - Dynamic Time Warping*, *Fuzzy K-Medoids - Euclidean*, *Fuzzy K-Medoids - Dynamic Time Warping*, untuk mengevaluasi kluster dan karakteristik pengelompokkan pada data Tingkat Pengangguran Terbuka pada 34 provinsi di Indonesia.

Data *time series* Tingkat Pengangguran Terbuka (TPT) memiliki karakteristik fluktuatif dan mengandung *outlier* pada periode tertentu. Kondisi tersebut membuat penggunaan metode *K-Means* menjadi tidak akurat karena sifat metode *K-Means* yang sangat sensitif terhadap *outlier* akibat perhitungan pusat klaster berbasis pada rata-rata. *K-Medoids* dipilih sebagai solusinya karena lebih tahan terhadap *outlier* dan bersifat non-parametrik, sehingga tidak memerlukan asumsi distribusi data. Metode *K-Medoids* bekerja berdasarkan ukuran jarak (*distance-based*) dengan menetapkan objek asli sebagai pusat *medoid* untuk memetakan wilayah secara tegas melalui pendekatan *hard clustering*. Pendekatan *hard clustering* pada *K-Medoids* bertujuan memetakan wilayah secara tegas ke dalam kelompok tertentu.

Pengelompokan secara tegas pada *hard clustering* memiliki keterbatasan dalam menangani objek yang berada di ambang batas antar-kelompok. Objek dengan pola data yang mirip sering kali terpaksa diklasifikasikan ke dalam satu kelompok secara mutlak, padahal secara karakteristik mungkin memiliki kedekatan dengan kelompok lain. Implementasi *Fuzzy K-Medoids* diterapkan untuk mengatasi kelemahan tersebut, khususnya pada data yang saling tumpang tindih (*overlapping*) dengan melakukan *clustering* secara *soft* yang tidak dapat dilakukan oleh *K-Medoids*. Penggunaan logika *fuzzy* membuat hasil *clustering* terlihat secara lebih detail, sehingga karakteristik wilayah di ambang batas antar-kelompok dapat teridentifikasi dengan presisi.

Pemilihan kedua metrik jarak ini dilakukan karena setiap metrik memiliki cara pandang yang berbeda dalam melihat data deret waktu TPT. Penggunaan jarak *Euclidean* bertujuan mengukur kedekatan berdasarkan besaran nilai pada titik waktu yang sama. *Dynamic Time Warping* digunakan untuk menangkap kemiripan

bentuk pola (*shape-based*) meskipun terdapat pergeseran waktu. Pada penelitian ini menggunakan dua jarak ini, sehingga dapat menentukan ukuran jarak mana yang paling konsisten dan mampu mencerminkan kesamaan pola pengangguran antarprovinsi secara lebih akurat sesuai dengan karakteristik data.

Evaluasi kluster dilakukan terhadap hasil kedua metode jarak dan klustering digunakan metrik *Silhouette*. Penggunaan *Silhouette* sebagai metode validasi kluster didasarkan pada kemampuannya mengevaluasi dua aspek sekaligus, yaitu kekompakan di dalam kelompok (*cohesion*) dan keterpisahan antar kelompok (*separation*). *Silhouette* memberikan informasi detail berupa nilai untuk setiap objek secara individu. Nilai tersebut mengidentifikasi secara presisi provinsi mana yang berada pada kelompok yang tepat serta provinsi yang berada di ambang batas antarkluster. *Silhouette* bersifat fleksibel terhadap penggunaan berbagai ukuran jarak, sehingga relevan untuk menggunakan hasil dari jarak *Euclidean* maupun *Dynamic Time Warping* pada data *time series*.

Penelitian ini diharapkan dapat menghasilkan pemetaan provinsi yang akurat sehingga nantinya dapat bermanfaat menjadi dasar kebijakan ketenagakerjaan yang lebih terarah untuk pemerintah pusat dan daerah dalam memahami karakteristik kelompok wilayah dengan pola pergerakan pengangguran yang serupa.

1.2 Rumusan Masalah

1. Bagaimana karakteristik jarak antarprovinsi di Indonesia berdasarkan data Tingkat Pengangguran Terbuka (TPT) dalam periode 2015 hingga 2025 menggunakan ukuran jarak *Euclidean* dan *Dynamic Time Warping* ?
2. Bagaimana pengelompokan provinsi menggunakan metode *K-Medoids* dan Fuzzy *K-Medoids* dengan jarak *Euclidean* dan *Dynamic Time Warping*?

3. Bagaimana jumlah klaster optimal berdasarkan *Silhouette Score* ?
4. Bagaimana karakteristik provinsi yang dihasilkan dari pengelompokan berdasarkan klaster optimum?

1.3 Batasan Masalah

1. Data yang digunakan adalah data sekunder yang bersumber dari BPS. Data yang digunakan hanya pada rentang waktu tahun 2015 hingga 2025. Selain itu, provinsi yang digunakan pada penelitian ini hanya mencakup 34 provinsi karena hasil pemekaran provinsi Papua (Papua Tengah, Papua Pegunungan, dan Papua Barat Daya) belum memiliki data TPT yang lengkap sejak tahun 2015 dan pada provinsi pemekaran tersebut sesuai Undang-Undang Nomor 14 Tahun 2022 baru dibentuk pada tahun 2022, sehingga sebelumnya masih tergabung dalam provinsi Papua dan Papua Barat. Ketidaklengkapan data menyebabkan provinsi hasil pemekaran tersebut tidak dapat digunakan sebagai unit analisis yang setara dalam penelitian *time series* ini.
2. Penelitian hanya menggunakan satu variabel, yaitu Tingkat Pengangguran Terbuka (TPT), sehingga faktor lain terkait pengangguran tidak dianalisis.
3. Evaluasi perbandingan metode klaster hanya dilakukan dengan *silhouette coefficient* berdasarkan metrik jarak yang sama dengan metode.
4. Banyaknya klaster yang digunakan dalam penelitian ini sebanyak $k = 2, 3, 4, \text{ dan } 5$.
5. Maksimum Iterasi (*max_iter*) untuk metode *Fuzzy K-Medoids* yang digunakan dalam penelitian ini adalah 50.

1.4 Tujuan Penelitian

1. Menganalisis karakteristik kemiripan antarprovinsi di Indonesia berdasarkan data Tingkat Pengangguran Terbuka (TPT) periode tahun 2015–2025 menggunakan metrik jarak *Euclidean* dan *Dynamic Time Warping* (DTW) untuk mengidentifikasi kedekatan pola data.
2. Mengimplementasikan metode pengelompokan *K-Medoids* dan *Fuzzy K-Medoids* menggunakan jarak *Euclidean* serta *Dynamic Time Warping* untuk memetakan pengelompokan provinsi berdasarkan karakteristik pola datanya.
3. Mengevaluasi hasil pengelompokan dengan nilai validasi *Silhouette Score* pada setiap metode untuk menentukan jumlah kluster dan ukuran jarak yang paling optimal.
4. Menginterpretasikan karakteristik wilayah pada setiap kluster yang terbentuk berdasarkan hasil pengelompokan optimal guna memberikan gambaran detail mengenai kondisi TPT antarprovinsi di Indonesia.