

# KLASTERISASI DATA PENGANGGURAN DI PULAU JAWA MENGUNAKAN ALGORITMA K-MEANS DALAM PENANGGULANGAN PENGANGGURAN TAHUN 2020-2023

Mutia Rachma<sup>1</sup>, Cahyono Budy Santoso<sup>2\*</sup>

<sup>1,2</sup>Sistem Informasi, Fakultas Teknologi dan Desain, Universitas Pembangunan Jaya, Tangerang Selatan, Indonesia

Email: <sup>1</sup>mutia.rachma@student.upj.ac.id, <sup>2\*</sup>cahyono.budy@upj.ac.id

(\* : coressponding author)

**Abstrak-**Pengangguran merupakan masalah utama dalam sektor perekonomian dan turut menimbulkan berbagai permasalahan sosial. Tingkat pengangguran ini muncul akibat dari ketidaksesuaian antara jumlah individu yang siap kerja dengan jumlah lapangan pekerjaan yang tersedia sehingga menimbulkan tantangan dalam penyerapan tenaga kerja. Penelitian ini bertujuan untuk mengidentifikasi dan mengelompokkan Kabupaten/kota di Pulau Jawa berdasarkan rata-rata Tingkat Pengangguran Terbuka (TPT) dan Tingkat Partisipasi Angkatan Kerja (TPAK) selama periode 2020-2023. Metode analisis yang digunakan adalah *clustering* dengan algoritma *K-Means*, dengan memanfaatkan data sekunder yang diolah berasal dari Badan Pusat Statistik (BPS) yang mencakup 119 Kabupaten/kota di enam provinsi di Pulau Jawa. Validasi jumlah kluster optimal dilakukan menggunakan *Silhouette score*, yang menunjukkan nilai tertinggi 0,55 menghasilkan dua kluster optimal. Hasil penelitian menunjukkan dua kelompok wilayah yang berbeda dalam karakteristik ketenagakerjaan. Kluster pertama terdiri dari 52 wilayah yang memiliki TPAK rendah dan TPT tinggi, mengindikasikan tantangan dalam penyerapan tenaga kerja yang lebih kompleks, terutama pada area urban atau pusat industri. Sebaliknya, kluster dua meliputi 67 wilayah yang memiliki TPAK tinggi dan TPT rendah, menunjukkan kondisi ketenagakerjaan yang relatif lebih stabil, seringkali di sektor pertanian atau pekerjaan *informal*. Analisis ini divisualisasikan menggunakan *scatter plot* dan *boxplot* untuk memperkuat interpretasi. Hasil klusterisasi ini diharapkan dapat menjadi acuan bagi pemerintah untuk menetapkan prioritas dan merumuskan kebijakan ketenagakerjaan yang lebih tepat sasaran sesuai dengan karakteristik masing-masing kluster wilayah di Pulau Jawa.

**Kata Kunci:** *K-Means*, Tingkat Pengangguran, Tingkat Partisipasi Angkatan Kerja, Klusterisasi, Pulau Jawa.

**Abstract-** *Unemployment poses a challenge within the economic sector, often triggering various social issues. This unemployment rate primarily stems from a mismatch between individuals ready to work and the available job opportunities, thereby creating challenges in labor absorption. This study aims to identify and classify regencies/cities in Java Island based on the average Open Unemployment Rate (TPT) and Labor Force Participation Rate (TPAK) during the period of 2020-2023. The analytical method employed is clustering using the K-Means, utilizing secondary data from Badan Pusat Statistik (BPS) that covers 119 cities across six provinces in Java Island. The optimal number of clusters was validated using the Silhouette score, which yielded the highest value of 0.55, indicating two optimal clusters. The results reveal two distinct regional groups with different employment characteristics. The first cluster consists of 52 regions, with low TPAK and high TPT, indicating more complex employment challenges typically found in urban or industrial areas. The second cluster includes 67 regions with high TPAK and low TPT, reflecting more stable employment conditions, often in agricultural/informal sectors. This analysis is visualized using scatter plots and boxplots to enhance interpretation. The result of clustering is expected to serve as a reference for the government in determining and formulating more targeted employment policies according to the specific characteristics of each regional cluster in Java Island.*

**Keywords:** *K-Means, Unemployment Rate, Labor Force Participation Rate, Clustering, Island Java.*

## 1. PENDAHULUAN

Pengangguran merupakan masalah utama dalam sektor perekonomian dan turut menimbulkan berbagai permasalahan sosial. Pengangguran merujuk pada individu yang mencari pekerjaan atau sedang mempersiapkan bisnis sebagai media sumber pendapatan, tingkat pengangguran ini terpengaruh oleh investasi domestik, asing, dan pertumbuhan ekonomi, sementara untuk inflasi tidak memiliki pengaruh yang berarti terhadap tingkat pengangguran di Indonesia [1]. Tingkat pengangguran ini muncul akibat dari ketidaksesuaian antara jumlah individu yang siap berkerja dengan jumlah lapangan pekerjaan yang tersedia sehingga menimbulkan tantangan dalam penyerapan tenaga kerja [2].

Salah satu pendekatan yang digunakan untuk memahami pola pengangguran, yaitu teknik *clustering* menggunakan algoritma *K-Means*. Penggunaan *clustering K-Means* telah banyak digunakan dalam berbagai studi untuk mengklasifikasikan data menjadi beberapa kelompok berdasarkan karakteristik tertentu. Muthmainnah et al, dalam penelitiannya menerapkan algoritma *K-Means* dengan tujuan untuk mengklusterisasi data pengangguran terbuka di Provinsi Jawa Barat, penelitian tersebut menghasilkan tiga kluster dengan karakteristik yang berbeda [3]. Untuk itu, penelitian ini relevan untuk melakukan mengklasifikasikan TPT dan TPAK diberbagai daerah khususnya Pulau Jawa guna mengindetifikasi karakteristik ketenagakerjaan [4][5]. Walaupun, studi sebelumnya sudah

membahas di wilayah tertentu, tepatnya di Provinsi Jawa barat, namun penerapan *clustering* di Pulau Jawa relatif jarang dibahas.

Permasalahan terkait pengangguran di Pulau Jawa menjadi tantangan besar, meski wilayah ini secara ekonomi tersentralisasi dengan lebih 58% perputaran ekonomi nasional terjadi di wilayah ini namun masih terdapat tingkat pengangguran yang tinggi [6]. Perbedaan tingkat partisipasi angkatan kerja dan tingkat pengangguran antar Kabupaten/kota menunjukkan ketimpangan dalam ketenagakerjaan yang perlu diperhatikan [7]. Beberapa Kabupaten/kota menunjukkan tingginya keterlibatan angkatan kerja diikuti oleh pengangguran terbuka bahwa peluang kerja belum mampu mengakomodasi seluruh tenaga kerja yang ada. Namun, rendahnya pengangguran terbuka yang dibarengi oleh rendahnya tingkat partisipasi kerja di suatu wilayah mengindikasikan bahwa hanya sebagian kecil usia produktif yang aktif secara ekonomi [8]. Oleh karena itu, keadaan ini perlu untuk melakukan pengelompokan wilayah guna menetapkan intervensi kebijakan yang tepat.

Metode analisis *clustering* dengan menggunakan algoritma *K-Means* untuk mengelompokkan Kabupaten/kota di Pulau Jawa berdasarkan Tingkat Pengangguran dan Tingkat Partisipasi Angkatan Kerja selama 2020-2023. Menggunakan *clustering* untuk mengklasifikasikan wilayah-wilayah dengan pola pengangguran dan partisipasi kerja dalam satu kluster. Wilayah-wilayah yang teridentifikasi dalam kluster yang memiliki tantangan lebih kompleks, dapat menjadi fokus utama pemerintah untuk merancang kebijakan yang lebih efektif serta prioritas penanganan yang sesuai [9]. Maliki et al, dalam penelitiannya menggunakan *clustering K-Means* untuk mengelompokkan provinsi di Indonesia berdasarkan TPT dalam dua periode waktu (2016-2018 dan 2019-2021), menghasilkan adanya perubahan signifikan antar kluster [10].

Penelitian ini diharapkan dapat menyajikan pemahaman terkait kondisi ketenagakerjaan dari Kabupaten/kota yang terdapat di Pulau Jawa kepada pemerintah, sehingga dapat mengkaji kebijakan yang lebih berorientasi pada target dalam menghadapi tantangan ketenagakerjaan di Pulau Jawa [11]. Pemahaman pola dan variasi tingkat pengangguran serta partisipasi angkatan kerja berbeda tiap wilayah, sehingga pemerintah lebih mudah untuk mengetahui wilayah yang memerlukan perhatian atau segera untuk ditangani [12].

Tujuan penelitian ini adalah untuk mengidentifikasi dan mengelompokkan Kabupaten/kota di Pulau Jawa berdasarkan rata-rata Tingkat Pengangguran Terbuka dan Tingkat Partisipasi Angkatan Kerja, serta memberikan informasi dalam mengkaji kebijakan-kebijakan terkait ketenagakerjaan dalam menanggulangi pengangguran dengan lebih tepat sasaran. Selain itu, memberikan gambaran menyeluruh bagaimana ketenagakerjaan di Pulau Jawa, sehingga hasil analisis ini dapat dijadikan landasan fundamental untuk mengkaji kebijakan-kebijakan yang relevan [2].

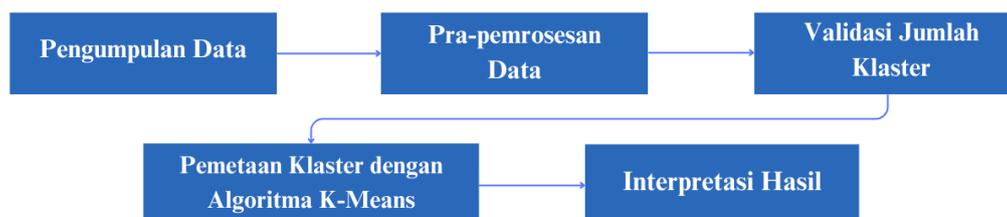
## 2. METODE PENELITIAN

### 2.1 Metode Data Mining

Pendekatan utama yang digunakan dalam penelitian ini adalah teknik *clustering* sebagai salah satu alat analisis dan penerapan algoritma *K-Means*. *K-Means* merupakan salah satu pendekatan untuk klusterisasi data yang bersifat *unsurvised learning*, yang mana bertujuan untuk mengorganisasi data menjadi beberapa kluster berdasarkan kemiripan karakteristik [13]. Karakteristik yang digunakan adalah rata-rata Tingkat Pengangguran Terbuka (TPT) dan Tingkat Partisipasi Angkatan Kerja (TPAK) dari masing-masing Kabupaten/kota di Pulau Jawa dari tahun 2020 hingga 2023. Proses ini tidak secara instan, melainkan melalui serangkaian langkah yang bertahap. Tahapan yang paling umum menentukan kluster yang akan dibentuk dan yang paling optimal dengan berbagai evaluasi [14]. *K-Means* ini digunakan untuk mengelompokkan wilayah berlandaskan pola ketenagakerjaan yang sebanding. Hasil klusterisasi diharapkan dapat memberikan gambaran yang jelas mengenai berbagai kelompok wilayah dengan berbagai tingkat partisipasi angkatan kerja dan keadaan pengangguran.

### 2.2 Tahapan penelitian

Penelitian ini melalui enam tahap untuk melakukan *clustering* data yaitu ditunjukkan dalam Gambar 1.



Gambar 1. Tahapan Penelitian

**a. Pengumpulan data**

Pada penelitian ini data yang digunakan yaitu 476 dataset mengenai data pengangguran dari enam provinsi per Kabupaten/kota di Pulau Jawa tahun 2020 hingga 2023 berdasarkan Tingkat Pengangguran Terbuka dan Tingkat Partisipasi Angkatan Kerja dari Badan Pusat Statistik (BPS). Menggunakan data sekunder memberikan ruang untuk efisiensi waktu dalam pengumpulan data, karena informasi yang diberikan sudah diolah terlebih dahulu oleh Badan Pusat Statistik yang mana merupakan lembaga yang terpercaya dari pemerintah [15].

**b. Pra-pemrosesan**

Setelah itu, tahap ini mencakup pembersihan data dan transformasi data untuk membersihkan data yang duplikasi atau data yang hilang serta menyesuaikan dengan format yang ada. Proses ini sangat penting untuk melakukan analisis lanjut dengan memastikan keakuratan data.

1) Pembersihan data

Pembersihan data adalah proses identifikasi dan penghapusan kesalahan pada data, yang memengaruhi kualitas hasil analisis. Proses ini mencakup verifikasi kualitas data dan koreksi kesalahan [16]. Pada penelitian ini tidak perlu melakukan pembersihan data dikarenakan data yang tersedia dari Badan Pusat Statistik (BPS) sudah memiliki validitas dan integrasi yang tinggi.

2) Transformasi data

Transformasi data yang dilakukan pada kolom Kabupaten/Kota bersifat kategorikal sehingga perlu diubah menjadi numerik menggunakan label *encoding* untuk keperluan analisis.

**c. Validasi Jumlah Kluster**

Penelitian ini menggunakan model *Silhouette score* untuk melakukan penentuan kluster optimal. *Silhouette* ini guna untuk mengukur tingkat akurasi tiap titik data yang berkelompok dalam kluster yang sama dan juga seberapa jauh titik-titik yang tersebar dari kluster lain. Di mana nilai yang lebih tinggi menunjukkan semakin baik data tersebut menyatu dengan klasternya sendiri dan semakin jelas terpisah dari kluster lain [17]. Metode *silhouette* ini dapat menunjukkan validitas kluster yang dibentuk dengan jelas, koefisien yang mendekati 1 menunjukkan pengelompokan yang baik [18].

**d. Pemetaan Kluster dengan Algoritma K-Means**

Algoritma *K-Means* merupakan metode populer untuk mengklasifikasikan data yang tidak memiliki struktur yang hierarki. Tujuan utamanya adalah untuk memetakan data ke dalam sejumlah kluster berdasarkan seberapa dekat terhadap *centroid* yang diukur secara iteratif. Metode ini populer dikarenakan memiliki kesederhanaan dan efisiensinya pada dataset kecil maupun besar. *K-Means* memilih titik acak sebagai pusat kluster, menghitung jarak antar *centroid* dan juga mengubah tempat *centroid* hingga konvergensi tercapai [19].

**e. Interpretasi Hasil**

Hasil dari proses pengklasifikasian ke dalam kluster, di interpretasikan untuk mengidentifikasi dan mengklasifikasikan Kabupaten/Kota di Pulau Jawa periode 2020-2023 berdasarkan Tingkat Pengangguran Terbuka dan Tingkat Partisipasi Angkatan Kerja sehingga dapat mengetahui kondisi ketenagakerjaan di Pulau Jawa.

**3. HASIL DAN PEMBAHASAN**

**3.1 Analisa Data**

Penelitian menggunakan data yang berasal dari 119 Kabupaten/kota dari masing-masing provinsi di pulau jawa 2020 hingga 2023 berdasarkan Tingkat Pengangguran Terbuka dan Tingkat Partisipasi Angkatan Kerja. Seluruh data TPT dan TPAK ini memiliki skala dan format yang konsisten sehingga tahap normalisasi tidak diperlukan untuk analisis lebih lanjut. Tabel 1 menyajikan rata-rata TPT dan TPAK sebagai hasil olahan yang digunakan dalam proses analisis.

Tabel 1. Rata-rata TPT dan TPAK

Kabupaten/Kota	Rata-rata TPT	Rata-rata TPAK
Kabupaten Lebak	8.4025	67.6975
Kabupaten Pandeglang	8.785	61.1475
Kabupaten Serang	10.8375	63.6975
Kabupaten Tangerang	9.235	65.8225
Kota Cilegon	9.5425	63.3225
Kota Serang	8.5725	64.7975
Kota Tangerang	7.905	64.74

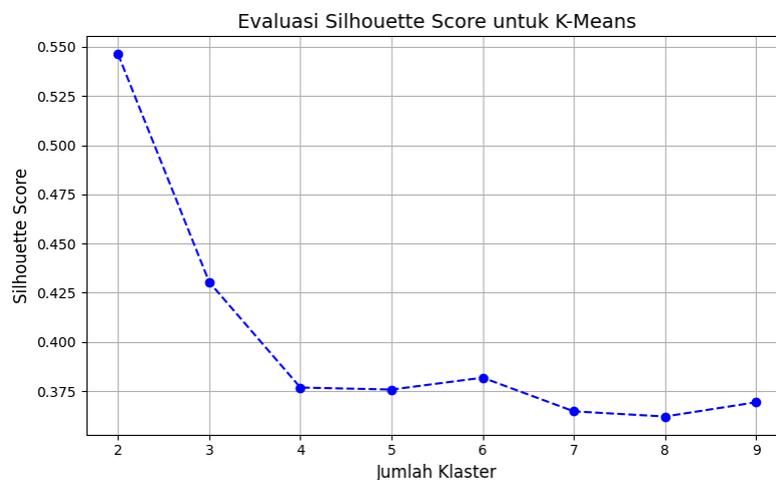
Kabupaten/Kota	Rata-rata TPT	Rata-rata TPAK
Kota Tangerang Selatan	7.37	61.055
Kabupaten Bantul	3.9625	73.885
Kabupaten Gunungkidul	2.1325	73.005
...	....	....
....	....	....
Kabupaten Trenggalek	4.3825	75.2025
Kabupaten Tuban	4.6075	73.4925
Kabupaten Tulungagung	5.455	73.02
Kota Batu	6.3625	74.1425
Kota Blitar	5.98	69.8825
Kota Kediri	5.255	68.7775
Kota Madiun	7.1775	67.91
Kota Malang	8.43	66.165
Kota Mojokerto	5.8475	69.3025
Kota Pasuruan	6.095	71.3525
Kota Probolinggo	5.5875	69.445
Kota Surabaya	8.4625	68.5175

Data yang digunakan diperoleh dari BPS berdasarkan Kabupaten/kota melalui salah satu tautan provinsi banten [https://bit.ly/BPS\\_Banten\\_2020](https://bit.ly/BPS_Banten_2020). Data serupa juga pada masing-masing provinsi lainnya di Pulau Jawa, yaitu DKI Jakarta, Jawa Barat, Jawa Tengah, DI Yogyakarta dan Jawa Timur. Tabel 1 merupakan data hasil olahan dari yang dipublikasikan oleh Badan Pusat Statistik (BPS). Nilai TPT dan TPAK rata-rata untuk setiap Kabupaten atau kota dihitung berdasarkan data yang tersedia dari masing-masing tahun. Data yang ditampilkan merupakan sampel data awal dan akhir, sementara seluruh data yang lengkap tetap digunakan dalam proses analisis kluster. Rata-rata TPT dan TPAK di setiap wilayah sangat berbeda seperti yang ditampilkan pada tabel. Tingkat pengangguran tinggi namun partisipasi angkatan kerja rendah ditemukan pada beberapa wilayah sementara yang lain sebaliknya. Kisaran TPT dan TPAK menunjukkan adanya perbedaan dalam keadaan ketenagakerjaan di antar wilayah.

### 3.2 Validasi Jumlah Kluster

Sebelum menentukan jumlah kluster ( $k$ ) yang digunakan dalam algoritma *K-Means*, evaluasi *Silhouette score* digunakan untuk menunjukkan seberapa tepat objek-objek yang tergolong dalam kluster yang sama dibandingkan kluster lain dan yang memiliki skor lebih tinggi menunjukkan pemisahan kluster yang terjadi lebih konsisten dan baik. Nilai *silhouette score* ini berkisar antara -1 dan 1. Evaluasi ini dengan mencoba variasi jumlah kluster mulai dari  $k=2$  hingga  $k=10$ . Hasil evaluasi yang dilakukan menggunakan *Silhouette score* ditampilkan pada Gambar 2.

Evaluasi menggunakan *Silhouette score* terlihat tertinggi pada  $k=2$  dengan nilai 0,545 atau 0,55, ini menunjukkan data yang terbagi sangat jelas menjadi dua kelompok yang berbeda. Kemudian saat ditambah menjadi tiga  $k=3$ , terlihat turun cukup tajam dengan nilai 0,43, ini menunjukkan data yang terbagi kurang jelas atau terjadi timpang tindih antar data. Pada  $k=4$  dan  $k=5$  dengan nilai yang relatif stabil namun cukup rendah, menunjukkan penambahan kluster tidak menyebabkan peningkatan kualitas segmentasi. Pada  $k=6$ , terjadi kenaikan nilai 0,38 yang dapat menunjukkan adanya sebagian kecil data yang mulai terpisah lebih jelas dari kluster lain.

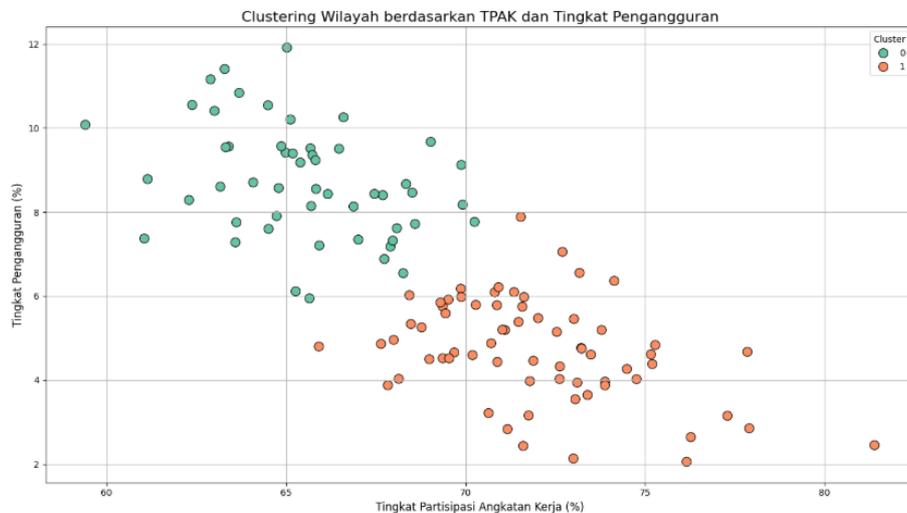


Gambar 2. Evaluasi *Silhouette Score*

Nilai kembali turun pada  $k=7$  dengan nilai 0,365 dan  $k=8$  dengan nilai 0,366, ini menunjukkan jumlah klaster yang terlalu banyak dapat membuat pembagian data menjadi terlalu spesifik. Terakhir, pada  $k=9$  dengan nilai yang sedikit naik 0,37, menunjukkan data yang tersebar tidak jelas. Maka, dua klaster yang optimal karena menghasilkan pemisahan yang paling jelas di antara data TPT dan TPAK Kabupaten/kota di Pulau Jawa. Pemilihan ini berdasarkan sisi realitas sosial dan ekonomi di lapangan. Klaster pertama terdiri dengan keadaan ketenagakerjaan relatif baik (TPT rendah dan TPAK tinggi) sedangkan klaster kedua dengan menghadapi keadaan yang lebih sulit biasanya terletak di wilayah perkotaan atau industri.

### 3.3 Pemetaan Klaster dengan Algoritma K-Means

Penyebaran data TPT dan TPAK menunjukkan perbedaan yang cukup jelas antara wilayah. Wilayah yang TPAK yang tinggi cenderung memiliki TPT yang rendah daripada wilayah dengan TPAK rendah atau sedang. Penyebaran ini, ada hubungan antara partisipasi kerja dan kapasitas serapan pasar tenaga kerja di masing-masing wilayah. Hasil klasterisasi dapat dilihat pada Gambar 3.



**Gambar 3.** Hasil Klasterisasi TPT dan TPAK

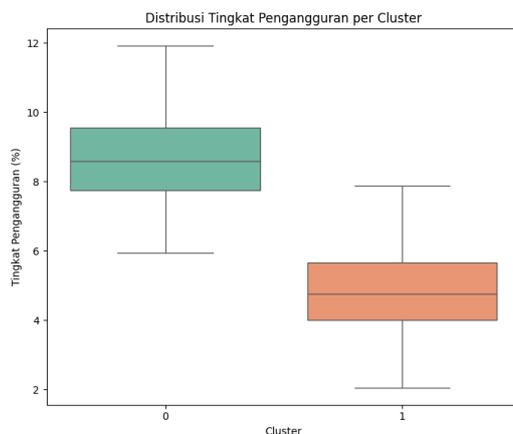
Gambar 3 menyajikan hasil dua klaster berdasarkan Tingkat pengangguran Terbuka (TPT) dan Tingkat Partisipasi Angkatan Kerja (TPAK). Klaster 0 yang berwarna hijau terdiri dari wilayah dengan TPT tinggi dan TPAK rendah (dibawah 70%) yang menunjukkan keadaan ketenagakerjaan yang mengalami tantangan dalam penyerapan tenaga kerja, sedangkan klaster 1 terdiri dari wilayah dengan TPT rendah dan TPAK tinggi yang menunjukkan keadaan ketenagakerjaan yang cenderung lebih produktif karena melibatkan banyak penduduk usia kerja dan mampu menyediakan lapangan kerja yang lebih baik. Visualisasi ini dapat dijadikan untuk mengidentifikasi wilayah yang harus diperhatikan lebih atau prioritas yang lebih terarah oleh pemerintah. Klasterisasi terhadap wilayah-wilayah dapat dilihat pada Tabel 2.

**Tabel 2.** Daftar Wilayah per Klaster

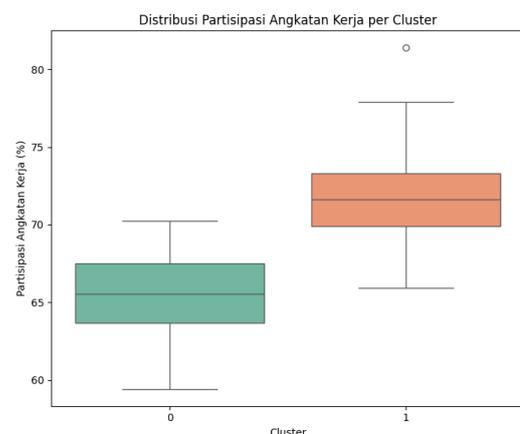
Klaster	Kabupaten/Kota
Klaster 0	Kabupaten Bandung, Kabupaten Bandung Barat, Kabupaten Bangkalan, Kabupaten Banyumas, Kabupaten Bekasi, Kabupaten Bogor, Kabupaten Brebes, Kabupaten Cilacap, Kabupaten Cirebon, Kabupaten Garut, Kabupaten Gresik, Kabupaten Indramayu, Kabupaten Karawang, Kabupaten Kuningan, Kabupaten Lebak, Kabupaten Pandeglang, Kabupaten Pematang, Kabupaten Purwakarta, Kabupaten Serang, Kabupaten Sidoarjo, Kabupaten Subang, Kabupaten Sukabumi, Kabupaten Sumedang, Kabupaten Tangerang, Kabupaten Tegal, Kepulauan Seribu, Kota Bandung, Kota Banjar, Kota Bekasi, Kota Bogor, Kota Cilegon, Kota Cimahi, Kota Cirebon, Kota Depok, Kota Jakarta Barat, Kota Jakarta Utara, Kota Madiun, Kota Magelang, Kota Malang, Kota Semarang, Kota Serang, Kota Sukabumi, Kota Surabaya, Kota Surakarta, Kota Tangerang, Kota Tangerang Selatan, Kota Tegal.
Klaster 1	Kabupaten Banjarnegara, Kabupaten Bantul, Kabupaten Banyuwangi, Kabupaten Batang, Kabupaten Blitar, Kabupaten Blora, Kabupaten Bojonegoro, Kabupaten Bondowoso, Kabupaten Boyolali, Kabupaten Ciamis, Kabupaten Demak, Kabupaten Grobogan, Kabupaten Gunungkidul, Kabupaten Jember, Kabupaten Jepara, Kabupaten Jombang, Kabupaten Karanganyar, Kabupaten Kebumen, Kabupaten Kediri, Kabupaten Kendal, Kabupaten Klaten, Kabupaten Kudus, Kabupaten Kulonprogo, Kabupaten Lamongan, Kabupaten Lumajang, Kabupaten Madiun, Kabupaten Maagelang, Kabupaten Magetan, Kabupaten Majalengka, Kabupaten Malang, Mojokerto, Kabupaten Nganjuk, Kabupaten Ngawi, Kabupaten Pacitan, Kabupaten Pamekasan, Kabupaten

Klaster	Kabupaten/Kota
	Pangandaran, Kabupaten Pasuruan, Kabupaten Pati, Kabupaten Pekalongan, Kabupaten Ponorogo, Kabupaten Probolinggo, Kabupaten Purbalingga, Kabupaten Purworejo, Kabupaten Remabang, Kabupaten Sampang, Kabupaten Semarang, Kabupaten Situbondo, Kabupaten Sleman, Kabupaten Sragen, Kabupaten Sukoharjo, Kabupaten Sumenep, Kabupaten Tasikmalaya, Kabupaten Temanggung, Kabupaten Trenggalek, Kabupaten Tuban, Kabupaten Tulungagung, Kabupaten Wonogiri, Kabupaten Wonosobo, Kota Batu, Kota Blitar, Kota Kediri, Kota Mojokerto, Kota Pasuruan, Kota Pekalongan, Kota Probolinggo, Kota Salatiga, Kota Yogyakarta.

Tabel di atas menyajikan daftar wilayah dari hasil pengelompokan ke dalam dua klaster. Pada klaster 0 terdiri dari 52 Kabupaten/Kota sedangkan klaster 1 mencakup 67 Kabupaten/kota. Secara keseluruhan, kedua klaster tersebut sebanyak 119 Kabupaten/kota. Kemudian, untuk memperkuat pada interpretasi hasil klasterisasi, diperlukan lebih dari melihat sebaran melalui *scatter plot*. Analisis distribusi statistik juga perlu dilakukan menggunakan *boxplot* yang mana menunjukkan karakteristik penyebaran data pada masing-masing klaster. *Boxplot* dapat dilihat pada Gambar 4a dan 4b.



**Gambar 4a.** Distribusi Tingkat Pengangguran Per Cluster



**Gambar 4b.** Distribusi Partisipasi Angkatan Kerja Per Cluster

Pada Gambar 4a terhadap distribusi tingkat pengangguran per cluster terlihat bahwa klaster 0 memiliki Tingkat pengangguran relatif tinggi dengan median 8,6% dan sebaran sebaran menilai hampir 12% yang menunjukkan tingkat pengangguran yang lebih serius. Sedangkan klaster 1 menunjukkan tingkat pengangguran rendah dengan median 5% dan distribusi yang lebih sempit yang menunjukkan tingkat pengangguran yang stabil di wilayah-wilayah tersebut. Kemudian Gambar 4b terhadap distribusi Tingkat partisipasi angkatan kerja per cluster terlihat bahwa klaster 0, TPAK yang lebih rendah dengan median 66% dan sebaran yang lebih sempit. Pada klaster 1, memiliki TPAK yang lebih tinggi dengan median 72% bahkan ada beberapa wilayah dengan TPAK melebihi 80% dengan terdapat *outlier*.

Perbedaan kedua yang terlihat jelas bahwa hasil klasterisasi ini berhasil membedakan dua kelompok wilayah berdasarkan karakteristik ketenagakerjaan. Dengan TPT tinggi dan TPAK rendah dan TPT rendah dan TPAK tinggi. Klaster 0 merepresentasikan wilayah yang menghadapi pada penyediaan lapangan kerja dan keterlibatan masyarakat dalam pasar kerja. Sementara, klaster 1 menunjukkan wilayah yang lumayan cukup maju secara ekonomi dengan rendahnya tingkat pengangguran dan tingginya partisipasi angkatan kerja.

### 3.4 Interpretasi Hasil

Hasil klasterisasi *K-Means* menunjukkan dua klaster wilayah Kabupaten/kota di Pulau Jawa berdasarkan Tingkat Pengangguran Terbuka (TPT) dan Tingkat Partisipasi Angkatan Kerja (TPAK) selama 2020-2023. Pada klaster 0 terdiri dari wilayah dengan Tingkat Partisipasi Angkatan Kerja cenderung rendah dan Tingkat Pengangguran Terbuka tinggi, dimana ini termasuk permasalahan yang cukup kompleks. Wilayah-wilayah yang tergolong pada klaster 0 cenderung pada sektor urban atau pusat industri, seperti Kota Jakarta, Kota Bandung dan Kota Bekasi. Ketidakseimbangan yang terjadi pada jumlah tenaga kerja dengan kesempatan atau lapangan pekerjaan menyebabkan terjadinya tingkat pengangguran yang melonjak meskipun fasilitas, pendidikan dan infrastruktur sudah tergolong cukup baik.

Sementara, pada klaster 1 terdiri dari wilayah dengan Tingkat Partisipasi Angkatan Kerja tinggi dan Tingkat Pengangguran Terbuka rendah yang mana keadaan ketenagakerjaan ini relatif stabil. Wilayah-wilayah ini berada dikawasan yang lebih mengandalkan sektor pertanian atau pekerjaan informal sebagai sumber pendapatannya. Masyarakat di wilayah ini banyak bekerja pada ladang, kebun atau sebagai pelaku usaha, namun juga pekerjaan

tersebut tidak menjamin dalam kestabilan ekonomi dalam jangka yang panjang. Misalnya, pada wilayah Kabupaten Pacitan, Kabupaten Gunung Kidul, dan Kabupaten Blora. Perbedaan dua klaster di Pulau Jawa menunjukkan adanya ketimpangan terkait keadaan ketenagakerjaan yang perlu ditangani dengan pendekatan kebijakan ketenakerjaan yang sesuai.

Dari penelitian ini dapat dijadikan sebagai referensi informasi berbasis data kepada pemerintah untuk mengkaji kebijakan-kebijakan mengenai ketenagakerjaan untuk menanggulangi pengangguran berdasarkan prioritas yang lebih tepat sasaran. Maka, pada klaster 0 membutuhkan penanganan lebih kompleks, dengan membutuhkan pelatihan vokasi, optimalisasi sektor industri dan lapangan pekerjaan dengan sektor industri yang lebih bervariasi. Selain itu, penting untuk memperkuat strategi pengembangan infrastruktur untuk menarik lebih banyak investor. Lalu, pada klaster 1 dapat difokuskan pada penguatan sektor *informal*, menerapkan peningkatan kualitas kerja dengan adanya program pelatihan untuk meningkatkan *skill* dari masing-masing individu, perlindungan terkait pekerjaan pada sektor pertanian atau *informal*. Penerapan program-program yang lebih ke arah untuk mendukung keberlanjutan ekonomi masyarakat, yaitu salah satunya pemberdayaan UMKM dengan didukung penuh oleh pemerintah.

#### 4. KESIMPULAN

Berdasarkan hasil penelitian ini, dapat disimpulkan bahwa penerapan algoritma K-Means berhasil mengklasterisasi 119 Kabupaten/kota di Pulau Jawa menjadi dua klaster berdasarkan rata-rata Tingkat Pengangguran Terbuka (TPT) dan Tingkat Partisipasi Angkatan Kerja (TPAK) selama periode 2020-2023. Klaster 0, yang terdiri dari 52 wilayah, menunjukkan kondisi ketenagakerjaan yang kompleks dengan TPAK yang cenderung rendah dan TPT yang tinggi. Wilayah-wilayah ini merupakan pusat urban atau industri, menghadapi masalah serius dalam penyerapan tenaga kerja meskipun infrastruktur dan fasilitas pendidikan relatif baik. Oleh karena itu, klaster ini memerlukan penanganan prioritas melalui penyediaan lapangan pekerjaan dengan sektor industri yang lebih bervariasi dan program pelatihan vokasi yang relevan. Sebaliknya, klaster 1, yang terdiri dari 67 wilayah, menunjukkan keadaan ketenagakerjaan yang relatif baik dengan TPAK yang tinggi dan TPT yang rendah. Wilayah-wilayah ini banyak bergantung pada sektor pertanian atau pekerjaan *informal*. Meskipun stabil, diperlukan program peningkatan *skill* individu dan dukungan penuh pemerintah untuk sektor *informal* guna menjamin keberlanjutan ekonomi jangka panjang. Model evaluasi menggunakan *Silhouette score* dengan nilai 0,55 yang cukup baik. Hasil klasterisasi juga didukung oleh visualisasi data melalui *scatter plot* dan *boxplot* yang sebagai analisis distribusi statistik. Hasil dalam penelitian ini, dapat dimanfaatkan sebagai pijakan oleh pemerintah untuk menetapkan prioritas terhadap wilayah yang sudah teridentifikasi berdasarkan klasternya untuk mengkaji kebijakan-kebijakan yang lebih tepat sasaran langsung kepada masyarakat.

#### DAFTAR PUSTAKA

- [1] B. Aflahza, D. E. Anggraini, M. Rusdi, Y. Yusnida, dan B. Benardin, "Analysis The Factors of Unemployment Rate in Indonesia on 1990-2022," *BICEMBA Bengkulu Int. Conf. Econ. Manag. Bus. Account.*, vol. 1, no. 1, hal. 38–42, 2023.
- [2] R. Maliqi, K. Falgenti, S. Priani, F. Fithri, M. Suherman, dan D. S. Nugraha, "Penerapan Metode K-Means Clustering pada Data Tingkat Pengangguran Terbuka Tahun 2016-2018 dan 2019-2021," *Comput. Sci.*, vol. 2, no. 2, hal. 109–116, 2022.
- [3] T. N. Muthmainnah, S. Indriyana, dan U. Enri, "Penerapan Algoritme K-Means Dalam Mengelompokkan Data Pengangguran Terbuka Di Provinsi Jawa Barat," *J. Inform. dan Rekayasa Perangkat Lunak*, vol. 5, no. 2, hal. 122, 2023.
- [4] E. Xiao, "Comprehensive K-Means Clustering," *J. Comput. Commun.*, vol. 12, no. 03, hal. 146–159, 2024.
- [5] N. Qona'ah, A. R. Devi, dan I. M. G. M. Dana, "Laboratory Clustering using K-Means, K-Medoids, and Model-Based Clustering," *Indones. J. Appl. Stat.*, vol. 3, no. 1, hal. 64, 2020.
- [6] A. R. dan N. Setuningsih, "Jokowi: 58 Persen Perputaran Uang Ada di Jawa, 17.000 Pulau yang Lain Dapat Apa? Artikel ini telah tayang di Kompas.com dengan judul 'Jokowi: 58 Persen Perputaran Uang Ada di Jawa, 17.000 Pulau yang Lain Dapat Apa?', Klik untuk baca: <https://nasional.kompas.com>," *Kompas.com*, 2023. [Daring]. Tersedia pada: <https://nasional.kompas.com/read/2023/03/17/08173901/jokowi-58-persen-perputaran-uang-ada-di-jawa-17000-pulau-yang-lain-dapat-apa>. [Diakses: 20-Mei-2025].
- [7] D. M. Tangke, "Analisis Upah Minimum Dan Pertumbuhan Ekonomi Sebagai Faktor Determinan Tingkat Pengangguran Terbuka Di Provinsi Maluku," *J. Ekon.*, vol. 13, no. 1, hal. 31, 2023.
- [8] U. N. Faizah dan N. Woyanti, "Analisis Pengaruh Pendidikan, Partisipasi Kerja, Dan Upah Minimum Terhadap Pengangguran Di Provinsi Banten Tahun 2011-2020," *BISECER (bus. Econ. Entrep.)*, vol. 6, no. 1, hal. 48, 2023.
- [9] S. Kusuma Arum, R. Astuti, dan F. Muhammad Basysyar, "Penerapan Algoritma K-Means Pada Dataset Pengangguran Terbuka Berdasarkan Pendidikan Di Provinsi Jawa Barat," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 8, no. 2, hal. 2221–2226, 2024.
- [10] R. Maliki, K. Falgenti, S. Priani, F. Fithri, M. Suherman, dan D. S. Nugraha, "Perbandingan Tingkat Pengangguran Terbuka Provinsi di Indonesia Berbasis Metode K-Means Clustering," *Comput. Sci.*, vol. 2, no. 2, hal. 109–116, 2022.
- [11] A. Bahauddin, A. Fatmawati, dan F. Permata Sari, "Analisis Clustering Provinsi Di Indonesia Berdasarkan Tingkat Kemiskinan Menggunakan Algoritma K-Means," *J. Manaj. Inform. dan Sist. Inf.*, vol. 4, no. 1, hal. 1–8, 2021.
- [12] M. F. Al Farizi *et al.*, "Pengelompokan Daerah di Jawa Timur Berbasis Indikator Kesejahteraan Masyarakat dengan Pendekatan Analisis Cluster Hierarki," *Inferensi*, vol. 6, no. 2, hal. 141, 2023.

- [13] N. Nurjanah, N. Suarna, dan W. Prihartono, “Implementasi K-Means Clustering Untuk Mengelompokan Tingkat Pengangguran,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 2, hal. 2462–2468, 2024.
- [14] B. Hakim, F. J. Kaunang, C. Susanto, J. Salim, and R. Indradjaja, “Implementasi Machine Learning Dalam Pengelompokan Musik Menggunakan Algoritma K-Means Clustering,” *IDEALIS : InDonEsiA journal Information System*, vol. 8, no. 1, pp. 74–83, 2025, doi: <https://doi.org/10.36080/idealis.v8i1.3357>.
- [15] T. Ambarsari, T. Rachmaningsih, dan H. Suryono, “Mengembangkan Diseminasi Official Statistics: Peluang dalam Perbaikan Pelayanan Publik (Developing Dissemination of Official statistics: Opportunities in Public Service Improvement),” *Semin. Nas. Off. Stat. 2021*, hal. 1–10, 2021.
- [16] F. Ridzuan dan W. M. N. Wan Zainon, “A review on data cleansing methods for big data,” *Procedia Comput. Sci.*, vol. 161, hal. 731–738, 2019.
- [17] R. W. Sembiring Brahmana, F. A. Mohammed, dan K. Chairuang, “Customer Segmentation Based on RFM Model Using K-Means, K-Medoids, and DBSCAN Methods,” *Lontar Komput. J. Ilm. Teknol. Inf.*, vol. 11, no. 1, hal. 32, 2020.
- [18] F. A. Husna, D. Purwitasari, B. A. Sidharta, D. A. Sihombing, A. Fahmi, dan M. H. Purnomo, “A Clustering Approach for Mapping Dengue Contingency Plan,” *Sci. J. Informatics*, vol. 9, no. 2, hal. 149–160, 2022.
- [19] M. Sitorus, M. Melisa, and Desty Herninda, “Driver Predicting Behavior Based on Accelerometer and Gyroscope Sensors with K-Means Algorithm Method,” *RIGGS Journal of Artificial Intelligence and Digital Business*, vol. 2, no. 1, pp. 12–17, 2023, doi: <https://doi.org/10.31004/riggs.v2i1.29>.