

ANALISIS CLUSTERING PENDUDUK BERDASARKAN KELOMPOK UMUR DENGAN *K-MEANS* DAN *HIERARCHICAL CLUSTERING* UNTUK PERENCANAAN DEMOGRAFI

Lutfiana Sinta Lestari

Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Tarumanagara,
Jln. Letjen S. Parman No. 1, Jakarta, 11440, Indonesia
E-mail: lutfiana.535210065@stu.untar.ac.id

ABSTRAK

Perencanaan demografi berperan penting dalam pembangunan wilayah, membantu pemerintah merancang kebijakan berdasarkan kebutuhan spesifik setiap kelompok umur. Namun, data demografi sering kali kompleks dan sulit ditafsirkan menggunakan metode tradisional. Penelitian ini menggunakan *K-Means* dan *Hierarchical Clustering* untuk mengelompokkan data penduduk berdasarkan kelompok umur. *K-Means* digunakan karena efisiensinya, sedangkan *Hierarchical Clustering* memberikan wawasan mendalam melalui dendrogram. Analisis menghasilkan tiga kluster utama: Kluster 2 (usia muda di bawah 40 tahun) mendominasi populasi (65%) dan menjadi prioritas untuk investasi di pendidikan dan lapangan kerja. Kluster 1 (usia 40–59 tahun) membutuhkan kebijakan produktivitas kerja, sedangkan Kluster 0 (usia lanjut di atas 60 tahun) memerlukan layanan kesehatan lansia dan program kesejahteraan sosial. Pemilihan tiga kluster sebagai solusi optimal didukung oleh evaluasi metrik seperti *Silhouette Score* dan *Calinski-Harabasz Score*. Hasil ini menunjukkan bahwa metode clustering efektif dalam menganalisis data demografi, memberikan wawasan yang mendalam untuk mendukung perencanaan berbasis data yang berkelanjutan. Penelitian ini juga berkontribusi pada literatur akademik mengenai analisis data demografi berbasis kluster.

Kata kunci—*Perencanaan Demografi, Analisis Kluster, K-Means, Hierarchical Clustering*

ABSTRACT

Demographic planning plays a vital role in regional development, helping governments design policies based on the specific needs of each age group. However, demographic data is often complex and challenging to interpret using traditional methods. This study employs K-Means and Hierarchical Clustering to group population data based on age groups. K-Means is utilized for its efficiency, while Hierarchical Clustering provides in-depth insights through dendrogram analysis. The analysis reveals three primary clusters: Cluster 2 (young individuals under 40 years) dominates the population (65%) and is prioritized for investments in education and employment opportunities. Cluster 1 (ages 40–59 years) requires policies aimed at enhancing work productivity, while Cluster 0 (elderly individuals over 60 years) necessitates elderly healthcare services and social welfare programs. The selection of three clusters as the optimal solution is supported by evaluation metrics such as the Silhouette Score and the Calinski-Harabasz Score. These results demonstrate that clustering methods are effective in analyzing demographic data, offering profound insights to support sustainable, data-driven planning. Furthermore, this research contributes to the academic literature on demographic data analysis using clustering techniques.

Keywords—*Demographic Planning, Cluster Analysis, K-Means, Hierarchical Clustering*

1. PENDAHULUAN

Perencanaan demografi adalah salah satu aspek yang sangat penting dalam pembangunan suatu wilayah [1]. Dengan memahami struktur penduduk, pemerintah dan pemangku kepentingan dapat merancang kebijakan yang sesuai untuk mendukung berbagai kebutuhan masyarakat, seperti pendidikan, kesehatan, ketenagakerjaan, hingga perencanaan infrastruktur. Salah satu komponen utama dalam perencanaan demografi adalah analisis kelompok umur, yang mencerminkan distribusi penduduk dalam berbagai segmen usia [2]. Informasi ini menjadi dasar bagi kebijakan yang

berorientasi pada kebutuhan spesifik setiap kelompok umur, seperti layanan pendidikan bagi anak-anak, program pelatihan kerja bagi kaum muda, dan pelayanan kesehatan bagi lansia.

Namun, tantangan utama dalam analisis data demografi adalah bagaimana mengelola dan menafsirkan data yang sering kali sangat besar dan kompleks[3]. Data distribusi penduduk tidak hanya bersifat multidimensi tetapi juga memiliki pola-pola tersembunyi yang sulit diidentifikasi melalui analisis tradisional. Dalam situasi ini, pendekatan berbasis data science, khususnya metode clustering, menawarkan solusi untuk mengelompokkan data ke dalam kategori yang bermakna berdasarkan karakteristik yang serupa. Metode ini memungkinkan analisis lebih mendalam untuk mengidentifikasi pola-pola spesifik dalam data, yang sebelumnya mungkin terlewatkan. Penelitian ini menggunakan dua teknik clustering, yaitu *K-Means* dan *Hierarchical Clustering*, untuk menganalisis data penduduk berdasarkan kelompok umur. *K-Means* dipilih karena algoritma ini terkenal akan kecepatannya dalam mengelompokkan data besar secara efisien[4]. Algoritma ini bekerja dengan cara membagi data ke dalam sejumlah cluster yang ditentukan sebelumnya berdasarkan jarak antara titik data dan pusat cluster. Di sisi lain, *Hierarchical Clustering* menawarkan pendekatan yang berbeda dengan membangun hierarki cluster yang divisualisasikan melalui dendrogram [5]. *Hierarchical Clustering* tidak hanya memberikan wawasan tentang struktur kelompok tetapi juga memungkinkan identifikasi hubungan antar *cluster*. Dengan menggabungkan kedua metode ini, penelitian bertujuan untuk memberikan analisis yang komprehensif terhadap pola distribusi penduduk berdasarkan kelompok umur. Analisis ini diharapkan mampu menghasilkan pengelompokan yang dapat digunakan untuk mendukung pengambilan keputusan berbasis data. Sebagai contoh, cluster yang terdiri dari dominasi anak-anak dan remaja dapat diarahkan pada kebijakan penguatan pendidikan dasar dan menengah, sementara cluster yang didominasi oleh lansia dapat menjadi dasar dalam perencanaan fasilitas kesehatan dan jaminan sosial. Hasil dari penelitian ini tidak hanya memberikan manfaat praktis dalam perencanaan demografi, tetapi juga berkontribusi pada pengembangan literatur akademik dalam bidang analisis data dan demografi. Dengan demikian, penelitian ini diharapkan dapat menjadi acuan bagi studi-studi serupa di masa depan dan mendukung pembangunan berbasis data yang berkelanjutan di era big data.

2. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan menerapkan metode clustering untuk menganalisis data penduduk berdasarkan kelompok umur. Dua teknik clustering yang digunakan dalam penelitian ini adalah *K-Means* dan *Hierarchical Clustering*, yang masing-masing memiliki keunggulan dalam mengelompokkan data dengan karakteristik yang berbeda. Penelitian ini dilakukan melalui beberapa tahapan berikut:

2.1. Pengumpulan Data

Pengumpulan dataset digunakan berasal dari website Badan Pusat Statistik. Data yang digunakan dalam penelitian ini mencakup distribusi penduduk berdasarkan kelompok umur untuk memahami pola distribusi penduduk laki-laki, perempuan, dan total penduduk dari tahun 2018 hingga 2023.

2.2. Pra-pemrosesan Data

Tahap ini bertujuan untuk memastikan kualitas data sebelum analisis dilakukan. Data mentah diolah untuk mengatasi masalah yang mungkin timbul, seperti data yang hilang (*missing values*) atau skala yang tidak seragam antar variabel. Teknik normalisasi digunakan untuk mengubah data ke dalam skala yang sama, sehingga mencegah dominasi variabel tertentu dalam proses clustering [6].

2.3. Implementasi *K-Means Clustering*

Algoritma *K-Means* diterapkan untuk mengelompokkan data penduduk ke dalam cluster berdasarkan kesamaan pola distribusi kelompok umur. Penentuan jumlah cluster optimal dilakukan

menggunakan Elbow Method, yaitu dengan mengevaluasi nilai inertia untuk berbagai jumlah cluster. Setelah jumlah cluster optimal diperoleh, algoritma K-Means digunakan untuk membagi data ke dalam cluster-cluster tersebut.

2.4. Implementasi *Hierarchical Clustering*

Hierarchical Clustering diterapkan sebagai metode pelengkap untuk memberikan pandangan yang lebih mendalam terhadap hubungan antar cluster. Metode ini menggunakan pendekatan *Ward's Linkage* untuk meminimalkan variansi dalam setiap cluster. Hasil analisis divisualisasikan dalam bentuk dendrogram, yang membantu mengidentifikasi jumlah cluster optimal berdasarkan pemotongan cabang dendrogram.

2.5. Analisis dan Interpretasi

Hasil clustering dari kedua metode dibandingkan untuk mengevaluasi konsistensi pola cluster yang dihasilkan. Setiap cluster dianalisis berdasarkan karakteristik kelompok umur yang dominan, sehingga dapat memberikan wawasan spesifik mengenai kebutuhan atau potensi kebijakan untuk setiap kelompok penduduk.

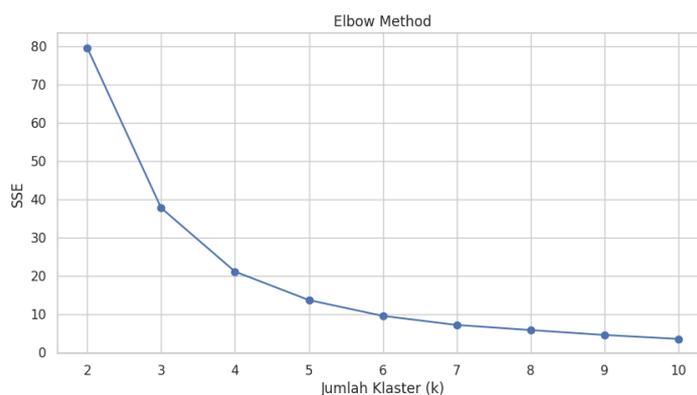
2.6. Visualisasi Hasil

Visualisasi hasil dilakukan untuk memperjelas interpretasi dan penyajian data [7]. Scatter plot digunakan untuk menampilkan hasil clustering K-Means, sementara dendrogram digunakan untuk menggambarkan struktur *Hierarchical Clustering*. Visualisasi ini bertujuan untuk membantu pembaca memahami pola distribusi penduduk dalam setiap cluster.

3. HASIL DAN PEMBAHASAN

Bagian ini, akan dipaparkan hasil dan pembahasan analisis clustering menggunakan metode K-Means dan Hierarchical Clustering terhadap data penduduk berdasarkan kelompok umur. Analisis ini bertujuan untuk mengidentifikasi pola distribusi penduduk yang dapat mendukung perencanaan demografi yang lebih efektif.

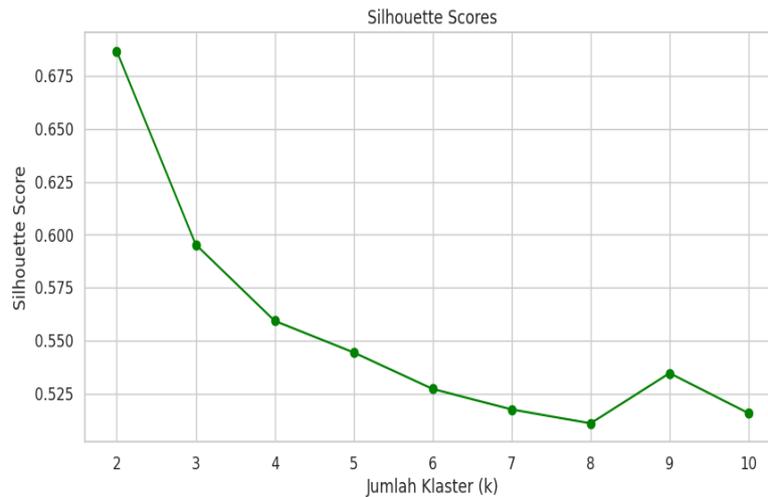
3.1. Clustering dengan K-Means



Gambar. 1. Elbow Method

Berdasarkan hasil analisis, grafik Elbow Method menunjukkan bahwa setelah cluster ke-3, penurunan nilai SSE menjadi lebih landai. Hal ini menunjukkan bahwa menambah jumlah cluster setelah $k=3$ tidak memberikan pengurangan signifikan pada variansi intra-cluster. Oleh karena itu, titik *elbow* pada grafik ini berada pada $k=3$, yang menjadikannya jumlah cluster optimal untuk analisis ini.

Grafik Elbow menampilkan hubungan antara jumlah kluster (sumbu X) dan nilai SSE (sumbu Y) [8]. Titik siku (*elbow point*) terlihat jelas pada $k=3$, menunjukkan bahwa tiga kluster adalah solusi terbaik dalam segmentasi data. Dengan memilih $k=3$, model clustering mampu mengelompokkan data secara efisien dengan keseimbangan yang baik antara kompleksitas model dan kualitas hasil clustering. Hasil dari Elbow Method mendukung penggunaan tiga kluster dalam analisis ini. Jumlah ini memberikan pemisahan yang baik antar kluster sekaligus menjaga interpretasi yang sederhana dan dapat diimplementasikan. Kluster optimal yang diidentifikasi akan digunakan untuk interpretasi lebih lanjut dalam konteks distribusi penduduk berdasarkan kelompok umur.



Gambar. 2. *Silhouette Scores*

Grafik *Silhouette Scores* memberikan wawasan penting dalam mengevaluasi kualitas clustering [9]. Nilai *Silhouette Score* berkisar antara -1 hingga 1, di mana nilai mendekati 1 menunjukkan bahwa data dalam kluster sangat mirip satu sama lain (homogenitas tinggi) dan terpisah dengan baik dari kluster lainnya [10]. Sebaliknya, nilai mendekati -1 menunjukkan bahwa data mungkin salah kluster.

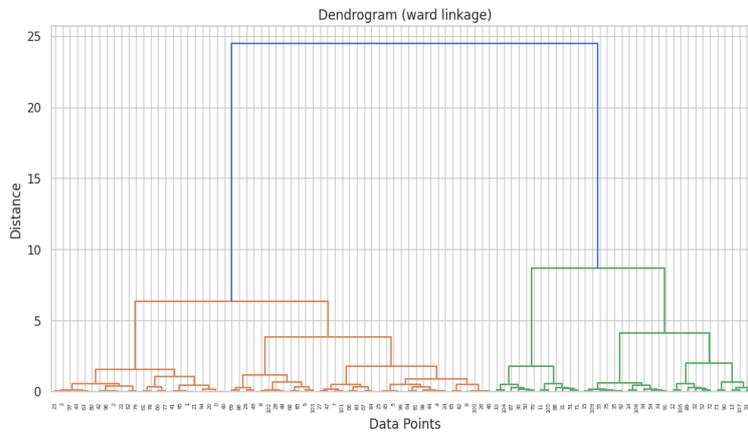
Berdasarkan hasil grafik, skor tertinggi tercapai pada $k=2$ dengan nilai sekitar 0.675. Hal ini menunjukkan bahwa dua kluster memberikan pemisahan paling jelas antar kluster sekaligus menjaga homogenitas dalam setiap kluster. Namun, ketika jumlah kluster ditambah menjadi $k=3$, *Silhouette Score* sedikit menurun menjadi sekitar 0.60. Meskipun skor ini lebih rendah dibandingkan $k=2$, nilainya masih cukup tinggi untuk dianggap sebagai hasil clustering yang baik. Interpretasi dan Keputusan

$K=2$ memberikan hasil clustering dengan kualitas terbaik dalam hal pemisahan dan homogenitas. Namun, dengan hanya dua kluster, hasilnya mungkin terlalu sederhana untuk mencerminkan kompleksitas distribusi data demografi.

$K=3$ menawarkan *trade-off* yang seimbang antara kualitas clustering dan jumlah kluster yang cukup untuk merepresentasikan variasi dalam data. Meski skor *Silhouette* sedikit lebih rendah dibandingkan $k=2$, pemilihan tiga kluster memberikan keunggulan dalam interpretasi yang lebih rinci dan relevan untuk tujuan analisis ini.

Secara keseluruhan, kluster $k=3$ dipilih sebagai solusi optimal dalam analisis ini. Pilihan ini mempertimbangkan keseimbangan antara performa metrik evaluasi dan kebutuhan untuk menghasilkan segmentasi data yang lebih informatif dan dapat digunakan dalam perencanaan demografi. Sementara $k=2$ menunjukkan kualitas teknis yang lebih tinggi, jumlah kluster yang lebih sedikit dapat membatasi interpretasi terhadap kompleksitas data. Dengan memilih $k=3$, analisis ini mampu memberikan wawasan yang lebih komprehensif sambil tetap menjaga kualitas clustering yang memadai.

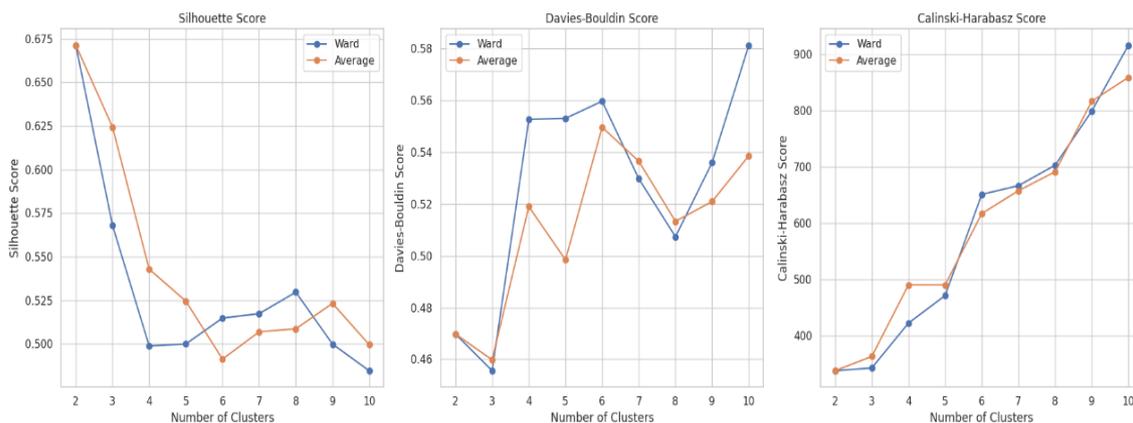
3.2. Evaluasi *Hierarchical Clustering*



Gambar. 3. Dendrogram Ward's linkage

Dendrogram di atas dihasilkan menggunakan metode Ward's Linkage, untuk meminimalkan variansi dalam setiap kluster yang terbentuk. Metode ini bekerja dengan menggabungkan dua kluster berdasarkan peningkatan terkecil dalam total variansi intra-kluster, sehingga menghasilkan kluster yang kompak dan homogen[12]. Pada dendrogram, sumbu X merepresentasikan titik data atau kelompok data, sedangkan sumbu Y menunjukkan tingkat perbedaan atau jarak (dissimilarity) antara kluster yang digabungkan. Penggabungan dimulai dari pasangan data dengan jarak terkecil hingga seluruh data tergabung dalam satu kluster di puncak dendrogram.

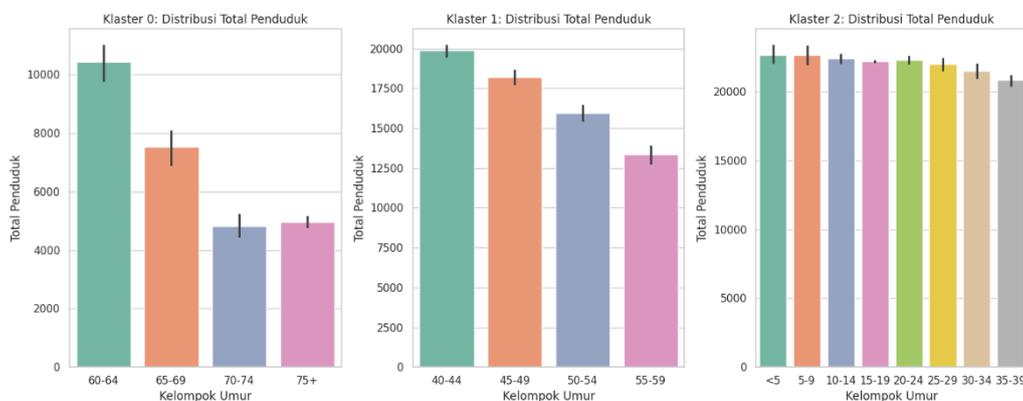
Berdasarkan grafik ini, pemotongan dendrogram pada tingkat jarak sekitar 10 menghasilkan tiga kluster utama, yang terlihat dari tiga kelompok cabang besar berwarna berbeda. Pemilihan tiga kluster ini memberikan pemisahan yang signifikan antar kluster dengan tingkat homogenitas yang baik dalam masing-masing kluster. Pendekatan Ward's Linkage sangat efektif dalam analisis demografi, karena memastikan bahwa setiap kluster merepresentasikan kelompok data dengan karakteristik yang mirip, sehingga memudahkan interpretasi dan mendukung segmentasi data yang bermakna [13].



Gambar. 4. Hasil Evaluasi Metrik

Berdasarkan hasil evaluasi metrik, metode Ward menunjukkan kinerja yang lebih unggul dibandingkan dengan metode Average [15]. Pada *Silhouette Score*, yang mengukur seberapa baik kluster saling terpisah dan serupa di dalamnya, metode Ward memberikan nilai yang lebih tinggi, menandakan kluster yang lebih baik. Begitu juga pada *Davies-Bouldin Score*, yang menilai sejauh mana kluster-kluster tersebut terpisah, metode Ward menghasilkan nilai yang lebih rendah, menunjukkan pemisahan kluster yang lebih jelas.

Sementara itu, Calinski-Harabasz Score, yang mengukur seberapa padat dan terpisah kluster, menunjukkan peningkatan seiring bertambahnya jumlah kluster, terutama pada metode Average. Namun, meskipun metode Average cenderung memiliki nilai yang lebih baik pada jumlah kluster yang lebih banyak, metode Ward tetap lebih unggul dalam hal pemisahan kluster yang lebih jelas dan konsisten, terutama pada jumlah kluster yang lebih sedikit (2 atau 3 kluster). Secara keseluruhan, jumlah kluster yang optimal untuk metode Ward adalah 2 atau 3 kluster, karena memberikan keseimbangan terbaik antara pemisahan kluster dan keseragaman antar anggota kluster. Meskipun metode Average memiliki hasil yang lebih baik pada jumlah kluster lebih tinggi, metode Ward dengan 2 atau 3 kluster adalah pilihan yang lebih optimal untuk mendapatkan hasil klusterisasi yang lebih baik. Sedangkan berdasarkan hasil evaluasi dengan tiga metrik klusterisasi, K-Means secara keseluruhan lebih baik daripada Hierarchical Clustering. K-Means memiliki *Silhouette Score* (0.595) dan *Calinski-Harabasz Score* (425.745) yang lebih tinggi, menunjukkan bahwa kluster yang dihasilkan lebih terpisah dan lebih bermakna. Meskipun Hierarchical Clustering sedikit lebih unggul dalam *Davies-Bouldin Score* (0.456), yang menunjukkan lebih sedikit tumpang tindih antar-kluster, perbandingan keseluruhan menunjukkan bahwa K-Means lebih unggul dalam kualitas klusterisasi.



Gambar. 5. Visualisasi data

Berdasarkan visualisasi data, ditemukan bahwa Klaster 2 mendominasi populasi dengan proporsi terbesar, yaitu 65%, diikuti oleh Klaster 1 dengan 24,8%, dan Klaster 0 dengan 10,2%. Klaster 2 didominasi oleh kelompok usia muda (di bawah 40 tahun), dengan distribusi gender yang sedikit lebih tinggi untuk laki-laki (51%) dibandingkan perempuan (49%). Populasi dalam kluster ini menunjukkan stabilitas dari tahun ke tahun, sehingga menjadi prioritas utama dalam alokasi sumber daya di bidang pendidikan, lapangan kerja, dan pengembangan teknologi. Sementara itu, Klaster 1 terdiri dari kelompok usia dewasa menengah (40-59 tahun) dengan distribusi gender yang seimbang. Klaster ini memiliki populasi yang cukup besar sehingga memerlukan kebijakan yang berfokus pada peningkatan produktivitas usia kerja, seperti pelatihan keterampilan dan insentif wirausaha. Di sisi lain, Klaster 0 terdiri dari populasi usia lanjut (60 tahun ke atas), dengan proporsi perempuan lebih tinggi (52%) dibanding laki-laki (48%). Oleh karena itu, kebijakan yang relevan untuk kluster ini mencakup peningkatan pelayanan kesehatan lansia, program dukungan sosial, dan pengadaan infrastruktur yang ramah lansia. Rekomendasi kebijakan bagi pemerintah dan lembaga demografi mencakup alokasi sumber daya berdasarkan kebutuhan spesifik setiap kluster. Untuk Klaster 2, investasi di bidang pendidikan, layanan kesehatan, dan infrastruktur publik menjadi prioritas utama guna mendukung populasi muda yang dominan. Klaster 1 memerlukan dukungan kebijakan ekonomi yang berorientasi pada pekerja usia produktif, seperti pelatihan kerja, insentif wirausaha, dan fleksibilitas pekerjaan. Sedangkan Klaster 0 membutuhkan perhatian khusus pada peningkatan layanan kesehatan lansia dan program kesejahteraan sosial, khususnya untuk perempuan yang lebih dominan dalam kelompok ini. Selain itu, perlu dilakukan analisis lebih lanjut terkait pengelompokan geografis, tren populasi, serta kaitannya dengan faktor ekonomi, untuk merancang kebijakan yang lebih efektif dan terarah sesuai kebutuhan masing-masing wilayah dan kelompok usia.

4.KESIMPULAN

Perencanaan demografi adalah elemen penting dalam pembangunan wilayah yang berkelanjutan, memungkinkan perumusan kebijakan yang sesuai dengan kebutuhan spesifik setiap kelompok umur. Penelitian ini menggunakan metode clustering, yaitu K-Means dan Hierarchical Clustering, untuk mengidentifikasi pola distribusi penduduk berdasarkan kelompok umur. Melalui pendekatan ini, ditemukan tiga klaster utama yang merepresentasikan karakteristik populasi yang berbeda. Klaster 2, yang mendominasi dengan proporsi 65%, didominasi oleh kelompok usia muda (di bawah 40 tahun) dengan distribusi gender hampir seimbang. Stabilitas populasi dalam klaster ini menunjukkan perlunya prioritas alokasi sumber daya untuk pendidikan, lapangan kerja, dan pengembangan teknologi. Klaster 1, yang terdiri dari usia dewasa menengah (40-59 tahun), memerlukan kebijakan yang mendukung produktivitas usia kerja, seperti pelatihan kerja dan insentif wirausaha. Klaster 0, yang terdiri dari populasi lansia (60 tahun ke atas), menunjukkan kebutuhan akan peningkatan layanan kesehatan, dukungan sosial, dan infrastruktur ramah lansia.

Hasil analisis menunjukkan bahwa K-Means memberikan hasil yang lebih baik dibandingkan Hierarchical Clustering dalam hal kualitas klasterisasi berdasarkan evaluasi metrik seperti Silhouette Score, Calinski-Harabasz Score, dan Davies-Bouldin Score. Pemilihan tiga klaster dianggap optimal karena memberikan keseimbangan antara interpretasi dan kualitas klasterisasi. Rekomendasi untuk pemerintah dan lembaga demografi meliputi

- 1). Alokasi Sumber Daya Berdasarkan Klaster: Pemerintah dan lembaga demografi disarankan untuk mengalokasikan sumber daya sesuai dengan kebutuhan spesifik tiap klaster, seperti pendidikan, kesehatan, atau pengembangan ekonomi.
- 2). Analisis Pengelompokan Geografis: Melakukan analisis lebih lanjut terhadap distribusi klaster secara geografis untuk memahami kebutuhan spesifik di berbagai wilayah.
- 3). Pemantauan Tren Populasi: Mengidentifikasi perubahan tren populasi dari waktu ke waktu untuk merancang kebijakan yang adaptif terhadap dinamika demografi.
- 4). Hubungan dengan Faktor Ekonomi: Mengintegrasikan data ekonomi, seperti tingkat pendapatan dan lapangan kerja, dengan analisis demografi untuk menghasilkan kebijakan yang lebih terarah.
- 5). Kontribusi untuk Pengembangan Literasi Demografi Berbasis Data: Penelitian ini memberikan kontribusi penting dalam pengembangan literatur akademik tentang analisis demografi berbasis data.
- 6). Acuan Pembangunan Berbasis Data: Hasil penelitian ini dapat menjadi dasar pengambilan keputusan untuk mendukung pembangunan yang lebih berkelanjutan dan berbasis data.

DAFTAR PUSTAKA

- [1] Sudianing, n. K., & seputra, k. A. (2019). Peran sistem informasi pemerintahan daerah dalam menunjang peningkatan kualitas perencanaan pembangunan daerah. *Locus*, 11(2), 112-133.
- [2] Sembiring, s. B. (2022). Analisis dinamika sosio-demografi sebagai acuan untuk mewujudkan rencana pemindahan ibu kota negara yang smart, green, beautiful dan sustainable. *Bappenas working papers*, 5(1), 120-137.
- [3] Pratama, a. S., sari, s. M., hj, m. F., badwi, m., & anshori, m. I. (2023). Pengaruh artificial intelligence, big data dan otomatisasi terhadap kinerja sdm di era digital. *Jurnal publikasi ilmu manajemen*, 2(4), 108-123.
- [4] Herdiana, o., maulani, s., & firdaus, e. A. (2021). Strategi pemasaran produk industri kreatif menggunakan algoritma k-means clustering berbasis particle swarm optimization. *Nuansa informatika*, 15(2), 1-13.
- [5] Wibowo, g. (2022). Segmentasi pelanggan berdasarkan recency frequency monetary (rfm) dan user event tracking menggunakan algoritma k-means clustering (budi luhur tesis).
- [6] Daniswara, a. A. A., & nuryana, i. K. D. (2023). Data preprocessing pola pada penilaian mahasiswa program profesi guru. *Journal of informatics and computer science (jinacs)*, 5(01), 97-100.

- [7] Sudipa, i. G. I., sarasvananda, i. B. G., prayitno, h., putra, i. N. T. A., darmawan, r., & wp, d. A. (2023). *Teknik visualisasi data*. Pt. Sonpedia publishing indonesia.
- [8] Painem, p., soetanto, h., & solichin, a. (2023). Analisis keberhasilan studi awal mahasiswa menggunakan klasterisasi k-means. *Faktor exacta*, 16(3).
- [9] Kurniawan, h. P., & farhatuaini, l. (2024). Identifikasi pola kepuasan mahasiswa terhadap proses pembelajaran menggunakan algoritma k-means clustering. *Jurnal informatika: jurnal pengembangan it*, 9(2), 164-172.
- [10] Yudistira, a., & andika, r. (2023). Pengelompokan data nilai siswa menggunakan metode k-means clustering. *Journal of artificial intelligence and technology information*, 1(1), 20-28.
- [11] Sadewo, m. G., eriza, a., windarto, a. P., & hartama, d. (2019, february). Algoritma k-means dalam mengelompokkan desa/kelurahan menurut keberadaan keluarga pengguna listrik dan sumber penerangan jalan utama berdasarkan provinsi. In *seminar nasional teknologi komputer & sains (sainteks)* (vol. 1, no. 1).
- [12] Indra, i., nur, n., iqram, m., & inayah, n. (2023). Perbandingan k-means dan hierarchical clustering dalam pengelompokan daerah beresiko stunting. *Inovtek polbeng-seri informatika*, 8(2), 356-367.
- [13] Sidebang, a. M. A. S. (2024). Pemetaan kemiskinan digital kabupaten/kota di sumatera utara menggunakan ward hierarchical clustering. *Journal of analytical research, statistics and computation*, 3(2), 23-39.
- [14] Sidebang, a. M. A. S. (2024). Pemetaan kemiskinan digital kabupaten/kota di sumatera utara menggunakan ward hierarchical clustering. *Journal of analytical research, statistics and computation*, 3(2), 23-39.
- [15] Fernandes, a. A. R. (2021). The use of ward linkage in cluster integration with a path analysis approach. *Open science index 15 2021*, 22, 92.
- [16] Pratama, k. A. Mapping the potential for food sector agricultural district/city in east java 2013 using cluster analysis method ward's.
- [17] Pramudiyanti, a. P., shafiro, m. F., & salim, l. A. (2024). K-means cluster analysis related to unmet need for family planning in banyuwangi, indonesia: a case study. *Journal of public health research & community health development*, 7(2).
- [18] Kastreva, p., & patarchanova, e. (2021). Creating spatial models of demographic processes using cluster analysis for demographic policy planning in bulgaria. *Journal of settlements & spatial planning*, 12(2).
- [19] Fuchs, m., & h pken, w. (2022). Clustering: hierarchical, k-means, dbscan. In *applied data science in tourism: interdisciplinary approaches, methodologies, and applications* (pp. 129-149). Cham: springer international publishing.
- [20] Sarbaini, s., saputri, w., & muttakin, f. (2022). Cluster analysis menggunakan algoritma fuzzy k-means untuk tingkat pengangguran di provinsi riau. *Jurnal teknologi dan manajemen industri terapan*, 1(2), 78-84.
- [21] Tendean, t., & purba, w. (2020). Analisis cluster provinsi indonesia berdasarkan produksi bahan pangan menggunakan algoritma k-means. *Jurnal sains dan teknologi*, 1(2), 5-11.
- [22] Astria, c., windarto, a. P., wanto, a., & irawan, e. (2019, august). Metode k-means pada pengelompokan wilayah pendistribusian listrik. In *seminar nasional sains dan teknologi informasi (sensasi)* (vol. 2, no. 1).
- [23] Astria, c., windarto, a. P., wanto, a., & irawan, e. (2019, august). Metode k-means pada pengelompokan wilayah pendistribusian listrik. In *seminar nasional sains dan teknologi informasi (sensasi)* (vol. 2, no. 1).
- [24] Zulyani, a. A., irawan, a. S. Y., & jamaludin, a. (2023). Penerapan data mining menggunakan algoritma k-means untuk menentukan tingkat vaksinasi pada kecamatan tambun selatan. *Innovative: journal of social science research*, 3(3), 7037-7050.
- [25] Sirait, n., & nababan, r. (2022). Implementation of data mining clustering astra vehicle insurance customers with the clustering method. *Jurnal manajemen, informatika, rekayasa perangkat lunak dan teknologi komunikasi*, 1(2), 73-85.