

ANALISIS *K-MEANS CLUSTERING* ANGKATAN KERJA BERDASARKAN JENIS KELAMIN, USIA, PENDIDIKAN, DAN PENDAPATAN PADA DATA SATUAN KERJA ANGKATAN NASIONAL TAHUN 2021

Sylviani Primaastuti Ananda^{1*}, Luluk Muthoharoh², Siectio Dicko Pratama³
^{1,2}Program Studi Sains Data, Fakultas Sains, Institut Teknologi Sumatera, Lampung
³Badan Pusat Statistik Lampung Utara, Lampung

*Penulis korespondensi: luluk.muthoharoh@sd.itera.ac.id

ABSTRAK

Pengelompokan karakteristik angkatan kerja di Indonesia merupakan langkah krusial untuk memahami dinamika pasar tenaga kerja yang terus berkembang, terutama dalam menghadapi tantangan ketimpangan sosial-ekonomi. Penelitian ini menggunakan metode *K-Means Clustering* dengan *Principal Component Analysis (PCA)* sebagai teknik reduksi data untuk menganalisis data Survei Angkatan Kerja Nasional (SAKERNAS) tahun 2021 yang diselenggarakan oleh Badan Pusat Statistik (BPS), dengan mempertimbangkan variabel jenis kelamin, usia, pendidikan terakhir, pekerjaan utama, dan pendapatan bulanan. Setelah melalui tahap *preprocessing* dan standarisasi data, dengan analisis *Elbow Method* mengidentifikasi 6 *cluster* sebagai jumlah optimal yang menggambarkan variasi sosial-ekonomi di Indonesia. Hasil penelitian menunjukkan bahwa pekerja dengan pendidikan lebih tinggi dan usia yang lebih tua cenderung bekerja di sektor formal dengan pendapatan yang lebih besar, sementara pekerja dengan pendidikan rendah dan usia lebih muda lebih banyak bekerja di sektor informal dengan pendapatan yang lebih kecil. Temuan ini menekankan peran penting pendidikan dalam memengaruhi sektor pekerjaan dan tingkat pendapatan, memberikan panduan bagi pemerintah dalam merancang kebijakan ketenagakerjaan yang lebih efektif. Dengan pemahaman yang lebih mendalam tentang pola pengelompokan ini, BPS dan pemerintah dapat mengembangkan program pelatihan vokasional dan pendidikan yang lebih sesuai untuk meningkatkan daya saing tenaga kerja serta mengurangi kesenjangan sosial-ekonomi di Indonesia.

Kata kunci: Badan Pusat Statistik, *Elbow Method*, *K-Means Clustering*, *Principal Component Analysis*, Survei Angkatan Kerja Nasional

1 PENDAHULUAN

Sebagai negara berkembang, Indonesia menghadapi tantangan besar dalam upaya meningkatkan kesejahteraan rakyatnya, khususnya dalam sektor ketenagakerjaan. Sumber daya manusia merupakan salah satu aset utama yang dapat mendorong pertumbuhan ekonomi di Indonesia. Namun, kualitas dan distribusi ketenagakerjaan di Indonesia masih menghadapi berbagai kendala yang ada. Berdasarkan hasil laporan Badan Pusat Statistik (BPS) tahun 2022, tingkat pengangguran terbuka di Indonesia masih mencapai 5,38%. Hal ini menunjukkan bahwa masih banyak sekali angkatan kerja yang belum terserap dalam lapangan kerja produktif, sehingga membuat kondisi menjadi semakin kompleks dengan adanya kesenjangan dalam akses terhadap

pendidikan, jenis kerjaan, dan pendapatan di berbagai wilayah tersebut (Aria, 2024). Oleh karena itu tingkat pendidikan dan keterampilan yang rendah akan menjadi hambatan bagi tenaga kerja di Indonesia untuk dapat bersaing di pasar kerja, baik tingkat nasional maupun global. Data menunjukkan bahwa mayoritas angkatan kerja di Indonesia masih berpendidikan rendah, dengan lebih dari 60% angkatan kerja hanya berpendidikan sampai tingkat sekolah menengah yang mengakibatkan rendahnya kualitas tenaga kerja dan produktivitas, yang akhirnya mempengaruhi tingkat pendapatan dan kesejahteraan (Arham & Akib, 2022). Untuk mendukung perumusan kebijakan yang tepat, Badan Pusat Statistik (BPS) berperan krusial dalam menyediakan data statistik yang *komprehensif*, salah satu survei yang dijalankan oleh BPS yaitu Survei Angkatan Kerja Nasional (SAKERNAS).

Badan Pusat Statistik (BPS) merupakan lembaga pemerintahan *non-dapertemen* yang memiliki tanggung jawab langsung di bawah Presiden dan berperan dalam mengumpulkan, mengolah, serta menyediakan informasi data statistik yang digunakan pemerintah dalam kebijakan dengan *komprehensif* (Fardi, 2022). Badan Pusat Statistik juga menyediakan informasi yang berkaitan langsung dengan ketenagakerjaan di Indonesia, salah satunya survei besar yang dilakukan oleh BPS secara rutin setiap tahunnya yaitu Survei Angkatan Kerja Nasional (SAKERNAS) yang merupakan sumber utama data ketenagakerjaan di Indonesia. Dengan tujuan untuk mendapatkan informasi menyeluruh terkait ketenagakerjaan tentang tenaga kerja, termasuk jumlah, karakteristik, dan kondisi kerja (Koniyyob & Zakariac, 2024). Data yang dihasilkan oleh Survei Angkatan Kerja Nasional menjadi dasar penting dalam proses perumusan kebijakan yang efektif dalam pemerintahan, khususnya yang berkaitan dengan ketenagakerjaan dan pembangunan ekonomi di Indonesia.

Seiring dengan berjalannya perubahan dinamis dalam lingkungan ekonomi global, struktur angkatan kerja di Indonesia mengalami transformasi yang cukup signifikan, terutama dengan munculnya sektor-sektor baru seperti ekonomi digital dan teknologi informasi yang menyebabkan variasi distribusi pendapatan dan membuka peluang kerja menjadi lebih luas (Arief Dirgantoro et al., 2014). Namun, dalam peningkatan jumlah angkatan kerja tidak selalu diimbangi oleh ketersediaan lapangan kerja yang memadai dan berkontribusi pada tingkat pengangguran tinggi. Untuk itu, dalam memahami pola karakteristik tenaga kerja dan untuk mendukung kebijakan ketenagakerjaan yang lebih efektif, penelitian yang dilakukan menggunakan metode *Clustering*. Khususnya dalam metode *K-Means* yang merupakan metode pengelompokan *non-hierarki* dan berguna dalam mengelompokkan data berdasarkan kesamaan karakteristiknya untuk membuat variabel menjadi satu kategori yang sama. Penelitian ini menggunakan metode *Clustering* pada data Survei Angkatan Kerja Nasional untuk analisis yang lebih mendalam terhadap hubungan antara kemiripan variabel-variabel data yang digunakan untuk menghasilkan rekomendasi kebijakan yang lebih tepat dalam mengatasi tantangan ketenagakerjaan di Indonesia (Haris Kurniawan et al., 2020). Mengingat besarnya data Survei Angkatan Kerja Nasional yang digunakan, metode *Principal Component Analysis (PCA)* diterapkan untuk mereduksi dimensi data, untuk meningkatkan efisiensi dan mempertajam akurasi dalam pengelompokan (Indhumathi & Sathiyabama, 2014). Sehingga dapat menghasilkan wawasan yang relevan bagi perumusan kebijakan ketenagakerjaan yang lebih relevan.

Penelitian ini berfokuskan menggunakan metode *K-Means Clustering* pada data Survei Angkatan Kerja Nasional, yaitu salah satu metode *unsupervised learning* yang efektif dalam analisis *Big*

Data, karena mampu mengidentifikasi pola tersembunyi berdasarkan kesamaan karakteristik tertentu (Yudistira & Adiputra, 2020). Dalam penelitian menggunakan data Survei Angkatan Kerja Nasional ini, variabel utama yang digunakan mencakup jenis kelamin, usia, pendidikan tertinggi, status pekerjaan dan pendapatan bulanan. Dengan besarnya data Survei Angkatan Kerja Nasional yang digunakan, metode *Principal Component Analysis (PCA)* yang diterapkan untuk mereduksi dimensi data secara efisien tanpa kehilangan informasi penting (Haris Kurniawan et al., 2020). Dengan tujuan untuk mengevaluasi keterkaitan antar variabel dalam setiap kelompok atau cluster yang terbentuk, untuk memahami apakah pendidikan memiliki pengaruh yang signifikan terhadap pekerjaan utama dan pendapatan yang menjadi isu sentral dalam kebijakan ketenagakerjaan di Indonesia. Serta melihat hasil kombinasi *PCA* dan *K-Means* yang memungkinkan pengelompokan yang lebih akurat dan efisien dalam menungkap pembaruan terkait variasi distribusi angkatan kerja, dalam pembaruan terkait variasi distribusi angkatan kerja dan pendidikan (Kumar et al., 2019).

2 METODE

2.1 Deskripsi Data

Dalam penelitian ini, menggunakan data Survei Angkatan Kerja Nasional (SAKERNAS) 2021, yang mencakup 104 kolom dan 3480 baris tentang kondisi tenaga kerja di Indonesia. Untuk analisis lebih lanjut, dipilih 5 variabel signifikan yaitu Jenis Kelamin, K6 (Usia), RA (Pendidikan Tertinggi), R12A (Status Pekerjaan Utama), dan R14A_UANG (Pendapatan Sebulan Terakhir). Selanjutnya dilakukan proses pembersihan data atau data *cleaning*, untuk menghapus data yang tidak lengkap, duplikat, atau tidak sesuai dengan kriteria analisis. Kemudian dataset akhir terdiri dari 5 kolom dan 1918 baris. Penelitian ini bertujuan menganalisis pola dan hubungan antar variabel tersebut untuk menggambarkan kondisi terkini tenaga kerja di Indonesia. Berikut 5 data data Sakernas teratas yang digunakan, dapat dilihat pada Tabel 1. Data Survei Angkatan Kerja Nasional.

Tabel 1. Data Survei Angkatan Kerja Nasional

No	Jenis Kelamin	K6	RA	R12A	R14A_UANG
1.	1	45	4	4	2400000
2.	2	37	3	7	4500000
3.	1	21	4	4	1000000
4.	1	50	4	4	2600000
5.	2	46	3	3	5811000

2.2 Elbow Method

Elbow Method adalah salah satu teknik yang digunakan dalam menentukan jumlah *cluster* yang paling tepat dalam algoritma *K-Means*. Metode ini bekerja dengan cara memvisualisasikan hasil *clustering* melalui grafik yang menggambarkan hubungan antara jumlah *cluster* dengan tingkat

kesalahan variansi dalam *cluster* (Ediyanto et al., 2013). *Elbow method* sering digunakan dalam algoritma *K-Means* untuk menentukan jumlah *cluster* yang paling optimal dengan cara membandingkan persentase pada perubahan variasi antar *cluster* (Kilic Depren et al., 2017). Pada penentuan jumlah *cluster* ini sangat penting dalam metode *clustering*, karena dengan pemilihan jumlah *cluster* yang tidak tepat akan menyebabkan hasil analisis yang kurang *representif*. Untuk itu, dengan menggunakan metode ini, jumlah dari *cluster* yang optimal dapat ditentukan pada saat perubahan penurunan nilai kesalahan atau *Sum of Square Error (SSE)* dari satu *cluster* ke *cluster* berikutnya akan melambat. Umumnya, dalam prose perhitungan *SSE*, jika semakin besar jumlah *cluster* yang ditambahkan, maka akan semakin kecil juga nilai *SSE*. Namun, penurunan terbesar dalam *SSE* akan terjadi pada iterasi awal dan ketika setelah jumlah *cluster* mencapai titik optimal maka penurunan *SSE* akan berkurang secara signifikan melalu titik penurunan terbesar terjadi di antara dua *cluster* pertama pada grafik (Pratama, 2018). Oleh karena itu, *Elbow method* akan sangat penting dalam analisis *K-Means* pada penelitian ini, karena akan membantu dalam memastikan jumlah *cluster* akan sesuai dengan struktur data yang ada, sehingga pada proses pengelompokan karakteristik dapat dilakukan lebih akurat.

2.3 Metode *K-Means*

Algoritma *K-Means* termasuk dalam kelompok metode pengelompokan data dalam kategori analisis cluster *non-hierarki*, yang di mana jumlah *cluster* yang akan dibentuk dan digunakan harus ditetapkan terlebih dahulu sebelum algoritma dijalankan. Hal ini berbeda dengan metode *hierarki*, yang memulai dengan menentukan sejumlah titik awal secara acak sebagai *cluster* yang disebut *centroid*. Algoritma ini kemudian akan melakukan iterasi berulang untuk mempartisi data ke dalam jumlah *cluster* yang telah ditentukan berdasarkan dengan kedekatan antara data dan *centroid* yang ditetapkan (Ni Putu Eka Merliana & Ernawati A.J.S, 2015). Tujuan utama dari *clustering* ini yaitu untuk mengidentifikasi struktur yang tersembunyi dalam data tanpa adanya label ataupun informasi sebelumnya terkait kategori ataupun kelompok yang ada, dengan memanfaatkan jika *Euclidean* dalam mengelompokkan objek-objek data ke dalam sejumlah kelompok (Arunkumar & Ganesh, 2019). Keunggulan dari algoritma ini terletak pada kemampuannya dalam menangani data berskala besar dengan memiliki efisiensi komputasi yang tinggi, sehingga sangat berguna dalam proses efisiensi dalam identifikasi pola distribusi yang tersembunyi di dalamnya (Syahfitri et al., 2023) seperti analisis ketenagakerjaan ataupun pasar tenaga kerja di wilayah yang luas. Secara matematis, algoritma *K-Means* menggunakan rumus :

$$\sum_{k=1}^K \sum_{x_i \in \pi_k} \|x_i - \mu_k\|^2 \quad (1)$$

Keterangan:

- K : jumlah *cluster*
- x_i : data ke-i
- π_k : himpunan data yang termasuk dalam *cluster k*
- μ_k : *centroid cluster k*
- $\|x_i - \mu_k\|^2$: jarak kuadrat *Euclidean* antara data x_i dan *centroid* μ_k

2.4 *Sillhouette Score*

Salah satu indikator evaluasi utama dari metode *K-Means* yaitu *Silhouette Score*, yang digunakan untuk menilai kualitas pengelompokan data dalam sebuah *cluster*. Metode ini dapat mengukur seberapa baik suatu titik data yang tergabung dalam *cluster* yang ditentukan, serta mengevaluasi seberapa jauh jarak dari *cluster* tetangga (Syahfitri et al., 2023). Nilai dari *Silhouette Score* berkisar antara -1 hingga 1, di mana skor yang mendekati 1 akan menunjukkan bahwa titik data tersebut berkelompok secara optimal dan baik dalam *cluster*. Sementara untuk nilai yang mendekati -1 mengindikasikan bahwa data data tersebut seharusnya lebih sesuai berada di *cluster* lain (Rianti et al., 2024). Oleh karena itu, evaluasi dengan hasil *Silhouette Score* tidak hanya membantu memahami pengelompokan, tetapi juga dapat memberikan wawasan dalam mengidentifikasi potensi perbaikan dalam permodelan untuk memastikan struktur data lebih akurat. Sehingga, dapat menghasilkan analisis yang relevan dalam pengelolaan data besar (Elly Muningsih, 2017), yang di mana hasil optimal akan berdampak signifikan terhadap keputusan yang diambil berdasarkan pola. Secara matematis, *Silhouette Score* dihitung menggunakan rumus :

$$s(x_i) = \frac{b(x_i) - a(x_i)}{\max \{b(x_i), a(x_i)\}} \quad (2)$$

Keterangan :

- $s(x_i)$: skor *silhouette* untuk data x_i
- x_i : anggota pada *cluster* π_k
- $a(x_i)$: jarak rata-rata x_i ke semua data lain di *cluster* yang sama
- $b(x_i)$: jarak rata-rata x_i ke semua data lain di *cluster* terdekat

$$d_l(x_i) = \frac{1}{|\pi_l|} \sum_{x_j \in \pi_l} d(x_i, x_j) \quad (3)$$

Keterangan:

- $d_l(x_i)$: rata-rata jarak antara titik x_i dengan semua titik data x_j dalam *cluster* l
- π_l : jumlah himpunan titik data yang ada di *cluster* l
- $|\pi_l|$: jumlah elemen dalam *cluster* l
- $d(x_i, x_j)$: jarak antara titik x_i dan x_j

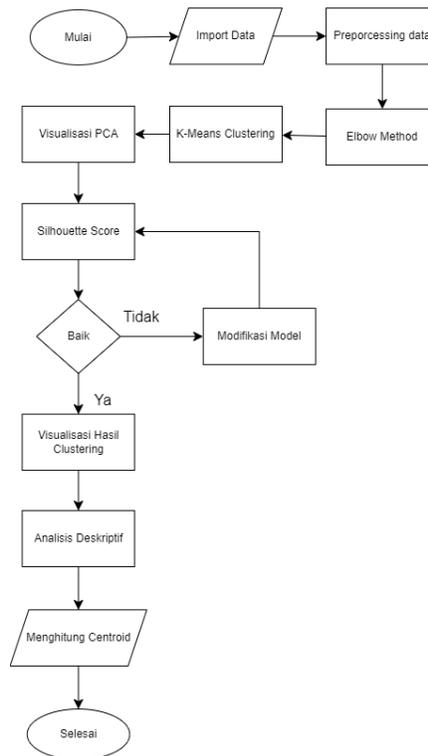
2.5 *Principal Component Analysis (PCA)*

Principal Component Analysis (PCA) merupakan salah satu teknik dalam statistik multivariat yang digunakan untuk memvisualisasikan hasil dari *K-Means Clustering* dan berfungsi sebagai metode reduksi dimensi pada *variable* (Rahayu & Mustakim, 2017). Teknik ini sangat berguna ketika jumlah variabel input yang digunakan terlalu besar dan berjumlah banyak, sehingga akan

menghambat kinerja komputasi akibat kompleksitas data (Hazmi Qastari et al., 2024). Tujuan utama dari *PCA* adalah menyederhanakan dimensi data tanpa kehilangan informasi yang signifikan, sehingga dapat memungkinkan analisis data menjadi lebih efisien dan cepat (Jamal et al., 2018). Penggunaan *PCA* dalam proses *K-Means Clustering* sendiri yaitu bertujuan untuk meningkatkan akurasi model prediksi, karena *PCA* dapat membantu mereduksi data yang banyak dan mempermudah analisis lanjutan dengan cara mengurangi jumlah variabel input dan menganalisis menjadi sederhana namun tetap dapat mempertahankan representasi utama dari data utama yang akan di analisis (Surono et al., 2023).

2.6 Diagram Alir Model

Pada Gambar 1. diagram alir pemodelan, menggambarkan proses pemodelan dalam penelitian. Dimulai dari *Import Data*, yaitu dengan menginputkan data yang diperoleh. Kemudian dilakukan *Preprocessing Data* untuk memastikan kualitas data, seperti pembersihan, penanganan nilai yang hilang, dan penyesuaian agar data siap digunakan. Langkah selanjutnya yaitu menggunakan metode *Elbow Method* untuk menentukan jumlah *cluster* optimal sebelum dilakukan proses *K-Means Clustering*.



Gambar 1. Diagram Alir Pemodelan

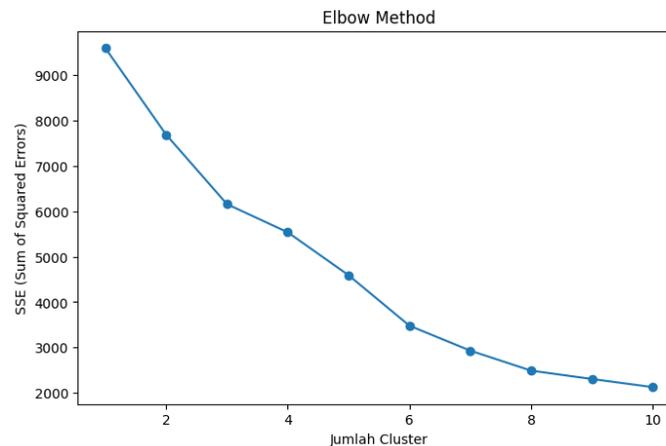
Hasil dari *cluster* yang didapatkan divisualisasikan menggunakan metode *Principal Component Analysis (PCA)* untuk menyederhanakan dimensi data dalam mempermudah interpretasi dan dievaluasi menggunakan *Sillhoutte Score* untuk mengukur kualitas pengelompokan. Jika hasilnya belum optimal, maka akan dilakukan *Modifikasi Model* hingga mencapai hasil yang maksimal. Setelah model dianggap baik, penelitian dilanjutkan dengan *Visualisasi Hasil Clustering*, *Analisis*

Deskriptif untuk memahami karakteristik setiap *cluster*, dan Perhitungan *Centroid* untuk menggambarkan ciri khas dari masing-masing *cluster* yang dihasilkan. Seluruh proses ini kemudian diakhiri pada tahap Selesai sebagai penutup dari rangkaian pemodelan.

3 HASIL DAN PEMBAHASAN

3.1 Hasil *Elbow Method*

Setelah data telah dilakukan standarisasi, kemudian dilakukan metode *Elbow Method* untuk menentukan jumlah *cluster* optimal dengan cara melihat persentase hasil perbandingan antara jumlah *cluster* yang membentuk siku pada suatu titik. Dari hasil Grafik *Elbow* pada Gambar 2. didapatkan bahwa titik optimal terjadi pada 6 *cluster*. Hal ini dapat dilihat bahwa bentuk siku mulai terlihat atau nilai *SSE* mulai menurun dengan lambat dan menjadi lebih datar secara signifikan setelah angka 6 pada sumbu x, yang artinya dengan menambahkan *cluster* lebih dari 6 tidak akan berpengaruh signifikan terhadap penurunan *SSE*. Kemudian untuk penerapan dalam metode *K-Means* selanjutnya akan menggunakan jumlah *cluster* dengan nilai 6. Dengan demikian, 6 *cluster* dipilih sebagai jumlah optimal yang dapat menggambarkan variasi sosial dan ekonomi dari angkatan kerja.

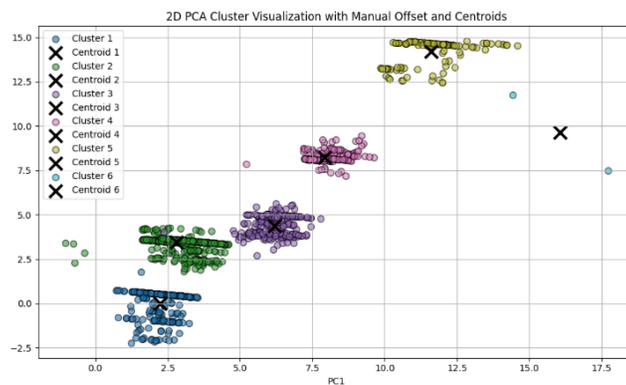


Gambar 2. Hasil *Elbow Method*

3.2 Hasil *K-Means Clustering*

Setelah menentukan jumlah *cluster* optimal, metode *K-Means Clustering* diterapkan pada hasil data yang telah di standarisasi dengan menghasilkan 6 *cluster*. Proses ini dilakukan dengan menggunakan metode *Principal Component Analysis (PCA)* untuk mereduksi dimensi data menjadi dua komponen utama, yaitu PC1 dan PC2. Hasil visualisasi yang diperoleh pada Gambar 2. menunjukkan pemisahan yang jelas antara kelompok-kelompok angkatan kerja di Indonesia berdasarkan dengan variabel sosial ekonomi yang telah di analisis. Kemudian setiap titik yang mewakili individu dalam data dikelompokkan berdasarkan hasil *K-Means Clustering*, dengan penerapan warna yang berbeda-beda untuk masing-masing *cluster* agar memungkinkan pemahaman yang lebih mudah terhadap struktur ekonomi yang ada. Selain itu, penggunaan *offset* pada posisi titik *cluster* dapat membantu dalam meningkatkan kejelasan dalam visualisasi, sehingga batas antar *cluster* yang berdekatan menjadi lebih jelas terbaca.

Analisis lebih lanjut terhadap *centroid* masing-masing *cluster*, yang ditandai dengan simbol 'x' berwarna hitam, dapat memberikan gambaran mengenai karakteristik rata-rata sosial ekonomi setiap kelompoknya. Dengan mengevaluasi posisi *centroid* dalam ruang *PCA*, dapat mengidentifikasi faktor utama yang membedakan *cluster* satu dengan yang lainnya. Hasil analisis menunjukkan bahwa 6 *cluster* angkatan kerja di Indonesia terbentuk berdasarkan dari karakteristik sosial ekonomi, yaitu jenis kelamin, usia, pendidikan terakhir, pekerjaan utama, dan pendapatan selama sebulan terakhir yang menghasilkan karakter unik tiap variabelnya. Sebagian besar dari kelompok, didominasi oleh individu dengan pekerjaan yang berpenghasilan tinggi dan tingkat pendidikan yang baik, sementara untuk kelompok lainnya lebih banyak terdiri dari pekerja muda dengan pendapatan yang lebih rendah. Hal ini mengindikasikan bahwa, jenis pekerjaan merupakan salah satu faktor yang signifikan dalam membedakan antar *cluster*, dengan perbedaan yang jelas antar individu yang bekerja di sektor formal dan informal. Hasil ini dapat dilihat secara jelas pada Gambar 2. yang menggambarkan hasil *K-Means Clustering* secara keseluruhan.



Gambar 3. Hasil *K-Means Clustering*

3.3 Hasil *Silhouette Score*

Silhouette Score merupakan metrik penting dalam menilai seberapa baik data ter-*cluster* dengan baik dan benar. *Silhouette Score* dapat berguna untuk memberikan gambaran terhadap seberapa dekat data dalam sebuah *cluster* dibandingkan dengan *cluster* lainnya. Semakin tinggi nilai *Silhouette Score* maka akan semakin baik pula hasil *clustering* nya. Dari hasil perhitungan pada data angkatan kerja, diperoleh hasil *Silhouette Score* dengan nilai 0.4158 yang dapat dilihat pada Gambar 3. Meskipun hasil nilai *Silhouette Score* telah cukup baik, namun terdapat grafik data yang berdempetan dan bertindih di antara beberapa *cluster*. Hal ini disebabkan oleh beberapa faktor, salah satunya yaitu disebabkan oleh jumlah data yang cukup besar dan bervariasi sehingga terdapat individu atau kelompok kecil yang memiliki karakteristik mirip dengan *cluster* lainnya.

Silhouette score: 0.4158219370754022

Gambar 4. Hasil *Silhouette Score*

4 KESIMPULAN

Berdasarkan analisis yang telah dilakukan, diperoleh kesimpulan bahwa penentuan jumlah *cluster* optimal dengan *Elbow Method* menghasilkan 6 *cluster* yang cukup *representatif* dalam menggambarkan variasi sosial ekonomi tenaga kerja di Indonesia berdasarkan jenis kelamin, usia, pendidikan terakhir, pekerjaan utama, dan pendapatan sebulan terakhir. Visualisasi pola sosial

ekonomi menggunakan *Principal Component Analysis (PCA)* berhasil menyederhanakan data ke dalam dua dimensi, menunjukkan pemisahan yang jelas antar *cluster* dan mempermudah analisis pola variasi sosial ekonomi. Analisis *centroid* mengidentifikasi karakteristik sosial ekonomi setiap *cluster*, termasuk distribusi jenis kelamin, usia, pendidikan, pekerjaan, dan pendapatan, serta menemukan individu-individu terdekat dari setiap *centroid* untuk memberikan pemahaman mendalam terkait kondisi angkatan kerja di Indonesia.

UCAPAN TERIMAKASIH

Puji syukur kehadirat Allah SWT atas rahmat dan karunia-Nya sehingga Laporan Kerja Praktik dengan judul "Analisis Pengelompokan Karakteristik Angkatan Kerja Berdasarkan Jenis Kelamin, Usia, Pendidikan, Pekerjaan, dan Pendapatan Menggunakan Metode K-Means Clustering pada Data Sakernas" dapat diselesaikan dengan baik. Penulis mengucapkan terima kasih kepada Ibu Lulu Muthoharoh, M.Si selaku dosen pembimbing, Bapak Sugaryadi, SE., MM selaku Kepala BPS Lampung Utara, Bapak Siectio Dicko Pratama, S.ST selaku pembimbing lapangan, seluruh staf BPS Lampung Utara, keluarga, serta teman-teman atas dukungan, arahan, dan bantuan yang diberikan selama proses kerja praktik dan penyusunan laporan ini. Penulis menyadari laporan ini masih memiliki kekurangan dan mengharapkan kritik serta saran yang membangun untuk perbaikan di masa mendatang.

DAFTAR PUSTAKA

- Arham, M. A., & Akib, F. H. Y. (2022). Supporting factors for labor productivity in Indonesia. *Jurnal Perspektif Pembiayaan Dan Pembangunan Daerah*, 10(5), 297–310. <https://doi.org/10.22437/ppd.v10i5.14447>
- Aria, R. R. (2024). Klasterisasi Angkatan Kerja Di Indonesia Berdasarkan Usia Menggunakan Metode Algoritma K-Means. *Inti Nusa Mandiri*, 18(2), 157–165. <https://doi.org/10.33480/inti.v18i2.5056>.
- Arief Dirgantoro, M., Mangkuprawira, S., Siregar, H., & Sinaga, B. . (2014). Dampak Kebijakan Desentralisasi Fiskal Terhadap Transformasi Ekonomi Di Provinsi Jawa Barat. *Jurnal Organisasi Dan Manajemen*, 5(1), 1–9. <https://doi.org/10.33830/jom.v5i1.205.2009>
- Arunkumar, & Ganesh, K. (2019). Performance and Analysis of Transmultiplexers. *Journal of Circuits, Systems, and Computers*, 28(1), 1–8. <https://doi.org/10.1142/S0218126619500099>
- Ediyanto, Mara, M. N., & Satyahadewi, N. (2013). Pengklasifikasian karakteristik dengan metode k-means cluster analysis. *Buletin Ilmiah Mat. Stat. Dan Terapannya (Bimaster)*, 02(2), 133–136.
- Elly Muningsih. (2017). Optimasi Jumlah Cluster K-Means Dengan Metode Elbow Untuk Pemetaan Pelanggan. *Prosiding Seminar Nasional Elinvo, September*, 105–114.
- Fardi, M. (2022). Sistem Informasi Perjalanan Dinas Pada Badan Pusat Statistik (BPS) Provinsi Lampung. *Ilmudata.Org*, 2(1), 1–11. <https://ilmudata.org/index.php/ilmudata/article/view/59>
- Haris Kurniawan, Sarjon Defit, & Sumijan. (2020). Data Mining Menggunakan Metode K-Means Clustering Untuk Menentukan Besaran Uang Kuliah Tunggal. *Journal of Applied Computer Science and Technology*, 1(2), 80–89. <https://doi.org/10.52158/jacost.v1i2.102>
- Hazmi Qastari, H. Soesanto, O. Sukmawaty, Y. (2022). K-Means Clustering dan Principal Component Analysis (PCA) Dalam Radial Basis Function Neural Network (RBFNN) Untuk Klasifikasi Data Multivariat, 4(2), 1-7.
- Indhumathi, R., & Sathiyabama, D. S. (2014). Reducing and Clustering high Dimensional Data

- through Principal Component Analysis. *International Journal of Computer Applications*, 11(8), 1–4. <https://doi.org/10.5120/1606-2158>
- Jamal, A. Handayani, A. Akbar Septiandri, A. Ripmiatin, E. Effendi, Y. (2018). Dimensionality Reduction using PCA and K-Means Clustering for Breast Cancer Prediction, 9(3), 192-201. <https://doi.org/10.24843/LKJITI.2018.v09.i03.p08>.
- Kılıç Depren, S., Aşkın, Ö. E., & Öz, E. (2017). Identifying the classification performances of educational data mining methods: A case study for TIMSS. *Kuram ve Uygulamada Eğitim Bilimleri*, 17(5), 1605–1623. <https://doi.org/10.12738/estp.2017.5.0634>
- Koniyob, F. C. O. M. H., & Zakariac, A. (2024). Perancangan Aplikasi Monitoring Kegiatan Survei Angkatan Kerja Nasional (SAKRNAS) di Badan Pusat Statistik Kota Gorontalo. *Diffusion Journal Of System And Information Tecnology*, 4(2), 93–102.
- Kumar, S. P., Bolla, S., & Anandan, R. (2019). A research on the importance of big data analytics in health care and government sectors. *International Journal of Engineering and Advanced Technology*, 9(1), 5471–5473. <https://doi.org/10.35940/ijeat.A3088.109119>
- Ni Putu Eka Merliana & Ernawati, A. J. S. (2015). Kajian Multi Disiplin Ilmu untuk Mewujudkan Poros Maritim dalam Pembangunan Ekonomi Berbasis Kesejahteraan Rakyat Analisa Penentuan Jumlah Cluster Terbaik Pada Metode K-Means. *Prosiding Seminar Nasional Multi Disiplin Ilmu & Call For Papers Unisbank (SENDI_U)*, 978–979. <https://www.unisbank.ac.id/ojs/index.php/sendu/article/view/3333>
- Pratama, S. A. (2018). *Analisis Faktor-Faktor Yang Mempengaruhi Tingkat Pengangguran Di Kabupaten Pelalawan*. Universitas Islam Riau.
- Rahayu, G & Mustakim. (2017). Principal Component Analysis untuk Dimensi Reduksi Data Clustering Sebagai Pemetaan Persentase Sertifikasi Guru di Indonesia, 201-208.
- Rianti, R., Andarsyah, R., & Awangga, R. M. (2024). Penerapan PCA dan Algoritma Clustering untuk Analisis Mutu Perguruan Tinggi di LLDIKTI Wilayah IV. *Nuansa Informatika*, 18(2), 67–77. <https://doi.org/10.25134/ilkom.v18i2.211>
- Surono, S. Wen Goh, K. Wou Onn, C. Marestiani, F. (2023). Developing an Optimized Recurrent Neural Network Model for Air Quality Prediction using K-Means Clustering and PCA Dimension Reduction, 6(2), 330-343.
- Syahfitri, N., Budianita, E., Nazir, A., & Afrianty, I. (2023). Pengelompokan Produk Berdasarkan Data Persediaan Barang Menggunakan Metode Elbow dan K-Medoid. *KLIK: Kajian Ilmiah Informatika Dan Komputer*, 4(3), 1668–1675. <https://doi.org/10.30865/klik.v4i3.1525>
- Yudistira, E. R., & Adiputra, I. M. P. (2020). Pengaruh Faktor Internal dan Eksternal Terhadap Harga Saham. *Jurnal Ilmiah Akuntansi Dan Humanika*, 10(2), 176. <https://doi.org/10.23887/jiah.v10i2.25862>