ANALISIS CLUSTERING DATA PENDUDUK BERDASARKAN USIA PRODUKTIF DI KECAMATAN BOGOR TIMUR MENGGUNAKAN K-MEANS CLUSTERING

Sahdan Dani Sena[1], Mulyati[2], Erniyati[3]
[1,2,3]Program Studi Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Pakuan
[1]sahdan.5725@gmail.com, [2]mulyati@unpak.ac.id, [3]neni_erniyati@unpak.ac.id

ABSTRAK

Penelitian ini bertujuan untuk menganalisis penduduk usia produktif di Kecamatan Bogor Timur menggunakan algoritma K-Means. Algoritma ini dipilih karena kemudahan penerapannya serta kecepatan dalam pemrosesan data. Metode penelitian yang digunakan bersifat kuantitatif, meliputi pengumpulan data, pemrosesan awal, penentuan jumlah cluster optimal, analisis, dan evaluasi. Pengelompokan dilakukan berdasarkan empat atribut utama, yaitu usia, jenis kelamin, pekerjaan, dan pendidikan. Jumlah cluster optimal dengan menggunakan metode Elbow adalah k=4. Hasil analisis masing-masing cluster didapatkan bahwa cluster 0, 1, dan 2 sebagian besar terdiri dari individu yang tidak atau belum bekerja dengan usia rata-rata yang bervariasi antara 17 hingga 38 tahun, sedangkan Cluster 3 didominasi oleh pekerja lepas dengan rata-rata usia adalah 41,5 tahun. Analisis lebih lanjut menunjukkan bahwa tingkat pendidikan sangat memengaruhi status pekerjaan. Mayoritas individu di Cluster 0 dan 1 hanya memiliki pendidikan SD atau SLTP. Sementara itu, cluster 2 terdiri dari individu yang masih dalam tahap pendidikan, dan Cluster 3 lebih banyak dihuni oleh mereka dengan pendidikan lebih tinggi yang bekerja secara mandiri. Hasil penelitian ini menunjukkan bahwa masih banyak penduduk usia produktif yang belum memiliki pekerjaan.

Kata Kunci : Clustering, K-Means, Metode Elbow, Usia Produktif,

ABSTRACT

This research aims to analyze the productive age population in East Bogor Sub-district using the K-Means algorithm. This algorithm was chosen because of its ease of application and speed in data processing. The research method used is quantitative, including data collection, initial processing, determining the optimal number of clusters, analysis, and evaluation. Clustering was done based on four main attributes, namely age, gender, occupation, and education. The optimal number of clusters using the Elbow method is k=4. The results of the analysis of each cluster found that clusters 0, 1, and 2 consisted mostly of individuals who were not or had not worked with an average age that varied from 17 to 38 years, while Cluster 3 was dominated by freelancers with an average age of 41.5 years. Further analysis shows that education level strongly influences employment status. The majority of individuals in Clusters 0 and 1 only have primary or secondary school education. Meanwhile, Cluster 2 consists of individuals who are still in the education stage, and Cluster 3 is more populated by those with higher education who work independently. The results of this study show that there are still many people of productive age who are unemployed.

***Keywords**: Clustering, Elbow Method, K-Means, Productive Age.*

PENDAHULUAN

Pertumbuhan jumlah penduduk usia produktif (15–64 tahun) di Indonesia merupakan potensi besar dalam mendorong pertumbuhan ekonomi nasional [1]. Namun, dalam realitasnya, tidak semua individu dalam rentang usia ini memiliki tingkat produktivitas yang sama. Faktor seperti pendidikan, keterampilan, status pekerjaan, dan tingkat kesejahteraan ekonomi dapat sangat bervariasi, sehingga diperlukan analisis lebih dalam untuk mengelompokkan usia produktif berdasarkan karakteristik tertentu guna mendukung strategi ekonomi dan sosial yang lebih efektif

Kecamatan Bogor Timur merupakan salah satu wilayah di Kota Bogor yang memiliki peran penting dalam pelayanan masyarakat serta pengelolaan data sosial ekonomi. Namun, salah satu tantangan utama yang dihadapi adalah minimnya analisis data yang dapat mengungkap pola serta faktor-faktor yang memengaruhi kependudukan. Data yang tersedia masih bersifat agregat dan statis, sehingga belum ada sistem yang mampu memetakan kelompok usia produktif maupun menggali informasi lebih mendalam dari data tersebut.

Salah satu solusi yang dapat diterapkan untuk mengatasi permasalahan ini adalah analisis clustering, yang memungkinkan pengelompokan data secara otomatis berdasarkan pola tersembunyi dalam data [2]. Metode K-Means dipilih karena mudah diimplementasikan [3] serta memiliki kecepatan komputasi yang tinggi [4]. Hal ini menjadi keuntungan tersendiri, terutama mengingat besarnya volume data kependudukan yang perlu dianalisis.

Berbagai penelitian telah menerapkan clustering K-Means dan memberikan hasil yang baik dalam banyak kasus [5] diantaranya dalam penelitian yang dilakukan oleh Parjito & Permata (2021) [6] serta Maulida & Haryadi (2024)[7] membahas bagaimana K-Means dapat digunakan untuk mengelompokkan data kemiskinan guna membantu perencanaan kebijakan sosial. Selain itu, penelitian oleh Afidah & Masrukan (2023)[8] dan Kuswanto et al. (2024) [9] menunjukkan bahwa metode ini bermanfaat dalam menganalisis pola migrasi penduduk serta tingkat pengangguran berdasarkan usia, yang berperan dalam perencanaan tenaga kerja dan pembangunan daerah. Penelitian yang dilakukan oleh Aria (2024)[10] dan Darmansah et al. (2024)[11] menunjukkan bahwa K-Means dapat digunakan untuk mengelompokkan angkatan kerja serta menganalisis tenaga kerja di sektor manufaktur, sehingga memberikan gambaran mengenai distribusi pekerja dan kesiapan industri.

Berdasarkan berbagai penelitian tersebut, dapat disimpulkan bahwa K-Means merupakan metode yang efektif dalam mengelompokkan data. Penerapan metode ini dapat membantu pengambilan keputusan berbasis data serta perumusan kebijakan yang lebih tepat sasaran.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif yang bertujuan untuk mengidentifikasi pola kelompok dalam data serta menentukan jumlah pengangguran pada usia produktif dengan algoritma K-Means Clustering. Pendekatan kuantitatif adalah metode penelitian yang mengumpulkan data dalam bentuk angka atau mengubah data kualitatif menjadi angka [12]. Langkah-langkah dalam penelitian ini dapat dilihat pada Gambar 1.

Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah data sekunder berupa data penduduk yang diperoleh dari Kantor kecamatan Bogor Timur, Dimana atributnya yang digunakan hanya, Usia, Jenis Kelamin, Pekerjaan, dan Pendidikan. dilakukan penyaringan data hanya usia produktif yang digunakan (15–64 tahun).

Preprossesing Data

Tahapan preprocessing yang dilakukan dalam penelitian ini yaitu: 1. Pengecekan missing Value [13] atau data yang hilang. 2. Mengubah data ke format yang lebih sesuai untuk analisis dalam hal ini Mengonversi kategorikal ke numerik. 3. Melakukan Standarisasi data dengan menggunakan metode Z-score, yaitu nilai

baku (Whendasmoro & Joseph,2022) yang mengubah setiap nilai sehingga memiliki rata-rata 0 dan standar deviasi 1.

Penentuan Jumlah Cluster Optimal

Menentukan jumlah cluster optimal dengan menggunakan metode Elbow[14]. Metode ini menentukan jumlah cluster optimal dengan cara menghitung nilai *Within-Cluster Sum of Squares* (WCSS) untuk beberapa nilai k (misalnya, k = 1, 2, 3, …). Nilai yang lebih kecil menunjukkan bahwa setiap klaster lebih konvergen [15]. Selanjutnya dilakukan visualisasi nilai k terhadap WCSS dalam sebuah grafik. Selanjutnya menentukan titik di mana penurunan WCSS mulai melambat secara signifikan, yang disebut titik "elbow". Titik ini dianggap sebagai jumlah cluster optimal, karena penambahan cluster selanjutnya tidak memberikan pengurangan WCSS yang berarti.

Analisis data dengan *K-Means Clustering*

Algoritma K-Means dikenal karena kesederhanaannya dalam penerapan, kecepatan proses, fleksibilitas, serta penggunaannya yang luas di berbagai bidang. Berikut adalah tahapan-tahapan dalam algoritma K-Means:

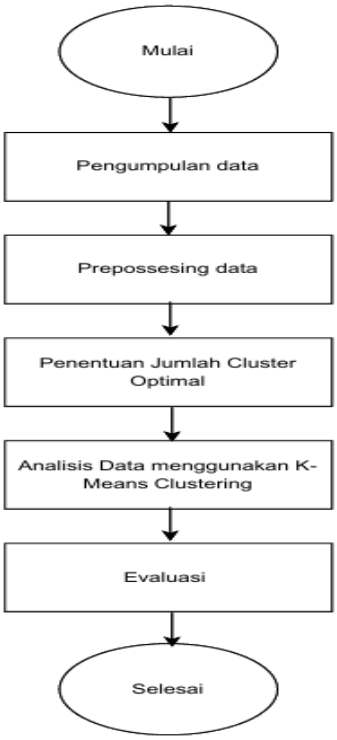
1. Inisialisasi titik pusat (centroid) awal: Pilih secara acak K titik data sebagai centroid awal untuk masing-masing cluster.
2. Perhitungan jarak data ke Setiap centroid: Hitung jarak antara setiap data dengan tiap centroid menggunakan rumus *Euclidean*. Rumus *Euclidean* untuk dua vector $x = (x_1, x_2, x_3, \dots x_n)$ dan $y = (y_1, y_2, y_3, \dots y_n)$

$$D(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \quad (1)$$

   dimana $D(x,y)$ adalah jarak antara objek *x* dan *y*, *n* : ukuran data, x_i dan *yi* merupakan nilai atau koordinat dari masing-masing titik pada dimensi ke-*i*.
3. Pengelompokan data: Setiap data akan dialokasikan ke cluster yang memiliki centroid terdekat berdasarkan perhitungan jarak.
4. Pembaharuan centroid: Untuk setiap cluster, hitung kembali titik pusat baru dengan cara mengambil rata-rata dari semua data yang termasuk dalam cluster tersebut. Proses ini memastikan centroid mewakili pusat data secara aktual.
5. Proses iterasi: Ulangi langkah perhitungan jarak dan pengelompokan data serta pembaharuan centroid hingga tidak terjadi perubahan signifikan pada posisi centroid atau keanggotaan data dalam cluster.

Evaluasi

Evaluasi hasil clustering sangat penting untuk memastikan bahwa pengelompokan data yang dilakukan sudah optimal dan sesuai dengan tujuan analisis. Dalam penelitian ini dilakukan analisis hasil pengelompokan berdasarkan variabel seperti usia, pendidikan, dan pekerjaan.



Gambar 1. *Flowchart* penelitian yang digunakan

HASIL DAN PEMBAHASAN

Data pada penelitian ini diperoleh dari instansi Kecamatan Bogor Timur, dengan total 66319 data individu. Data mencakup informasi seperti Nama, Status Dalam Keluarga, Usia, Jenis Kelamin, Pekerjaan, dan Pendidikan. Adapun tampilan Sebagian dataset dapat dilihat pada Gambar 2:

	A	B	C	D	E	F	G	H	I
1	**NAMA**	**KELUARGA**	**ALAMAT**	**KECAMATAN**	**KELURAHAN**	**USIA**	**JENIS KELAMIN**	**PEKERJAAN**	**PENDIDIKAN**
2	HASANUDIN	Kepala Keluarga	KP.BATAKAL	SELATAN	BATU TULIS	52	Laki-laki	Pekerja Lepas	SLTP/Sederajat
3	ERNAWATI	Istri/Suami	KP.BATAKAL	SELATAN	BATU TULIS	52	Perempuan	Bekerja	SD/Sederajat
4	SUSILAWATI	Anak	KP.BATAKAL	SELATAN	BATU TULIS	12	Perempuan	Bekerja	SD/Sederajat
5	RAMDANI	Anak	KP.BATAKAL	SELATAN	BATU TULIS	19	Laki-laki	Bekerja	SLTA/Sederajat
6	ARMAN	Anak	KP.BATAKAL	SELATAN	BATU TULIS	22	Laki-laki	Wiraswasta	SLTA/Sederajat
7	ARI SUTIAWAN	Anak	KP.BATAKAL	SELATAN	BATU TULIS	28	Laki-laki	Pekerja Lepas	SLTP/Sederajat
8	SRI WULAN	Kepala Keluarga	WARNA 04	SELATAN	BATU TULIS	59	Perempuan	Pegawai Swasta	SD/Sederajat
9	YULIYANTO	Anak	WARNA 04	SELATAN	BATU TULIS	22	Laki-laki	Pegawai Swasta	SD/Sederajat
10	KUSUMAH	Anak	WARNA 04	SELATAN	BATU TULIS	29	Laki-laki	Pegawai Swasta	SLTA/Sederajat
11	SETIAWAN	Anak	WARNA 04	SELATAN	BATU TULIS	33	Laki-laki	Bekerja	SLTA/Sederajat
12	DENI MULYADI	Kepala Keluarga	CEMPAKA	SELATAN	BATU TULIS	48	Laki-laki	Pekerja Lepas	SLTA/Sederajat
13	TRI YULIANTI	Istri/Suami	CEMPAKA	SELATAN	BATU TULIS	52	Perempuan	Bekerja	SLTA/Sederajat
14	REZI	Anak	CEMPAKA	SELATAN	BATU TULIS	14	Laki-laki	Bekerja	SD/Sederajat
15	RAEZA	Anak	CEMPAKA	SELATAN	BATU TULIS	14	Laki-laki	Bekerja	SD/Sederajat
16	RIZKI MAULANA	Anak	CEMPAKA	SELATAN	BATU TULIS	18	Laki-laki	Bekerja	SLTP/Sederajat
17	RAHMAWATI	Anak	CEMPAKA	SELATAN	BATU TULIS	24	Perempuan	Pegawai Swasta	SLTP/Sederajat
18	RUDI KARTIWAN	Kepala Keluarga	WARNA IV	SELATAN	BATU TULIS	55	Laki-laki	Pekerja Lepas	SD/Sederajat
19	SURYANIH	Istri/Suami	WARNA IV	SELATAN	BATU TULIS	53	Perempuan	Pegawai Swasta	SD/Sederajat

Gambar 2. Tampilan Sebagian Dataset

Dataset yang sudah dikumpulkan selanjutnya dilakukan filtering atribut untuk menyaring dan memilih variabel-variabel yang relevan serta berpengaruh signifikan terhadap analisis. Atribut yang digunakan dalam penelitian ini adalah Usia, Jenis Kelamin, Pekerjaan, dan Pendidikan. Atribut Usia dilakukan lagi penyaringan data hanya usia produktif yang digunakan yaitu rentang usia antara

15–64 tahun sehingga total data yang digunakan berjumlah 46582 data. Gambar 3 berikut hasil filtering data dengan menggunakan *Pemrongraman Phyton*.

	USIA	JENIS KELAMIN	PEKERJAAN	PENDIDIKAN
0	52	Laki-laki	Pekerja Lepas	Tamat SLTP/Sederajat
1	52	Perempuan	Tidak/Belum Bekerja	Tamat SD/Sederajat
3	19	Laki-laki	Tidak/Belum Bekerja	Masih SLTA/Sederajat
4	22	Laki-laki	Wiraswasta	Tamat SLTA/Sederajat
5	28	Laki-laki	Pekerja Lepas	Tamat SLTP/Sederajat
...	...	...	...	...
66314	57	Laki-laki	Wiraswasta	Tamat SD/Sederajat
66315	49	Perempuan	Tidak/Belum Bekerja	Tamat SD/Sederajat
66316	15	Perempuan	Tidak/Belum Bekerja	Masih SLTP/Sederajat
66317	19	Perempuan	Tidak/Belum Bekerja	Tamat SLTP/Sederajat
66318	54	Perempuan	Tidak/Belum Bekerja	Tamat SD/Sederajat

Gambar 3. Hasil Filtering data menggunakan atribut yang relevan

Tahap awal yang dilakukan adalah dengan melakukan pengecekan data apakah terdapat data yang hilang atau *missing value*. Selanjutnya, Atribut Jenis Kelamin, Pekerjaan dan Pendidikan dikonversi dari kategorik menjadi numerik menggunakan pengkodean phyton dapat dilihat pada Gambar 4.

```
In [4]: from sklearn.preprocessing import LabelEncoder

        # Inisialisasi LabelEncoder
        label_encoders = {}
        categorical_columns = ["JENIS KELAMIN", "PEKERJAAN", "PENDIDIKAN"]

        # Melakukan encoding pada setiap kolom kategorikal
        df_encoded = df_filtered.copy()
        for col in categorical_columns:
            le = LabelEncoder()
            df_encoded[col] = le.fit_transform(df_encoded[col])
            label_encoders[col] = le  # Simpan encoder untuk referensi

        # Menampilkan beberapa baris pertama setelah encoding
        df_encoded.head()
```

Out[4]:

	USIA	JENIS KELAMIN	PEKERJAAN	PENDIDIKAN
0	52	0	5	7
1	52	1	8	5
3	19	0	8	2
4	22	0	9	6
5	28	0	5	7

Gambar 4. Hasil dikonversi dari kategorik menjadi numerik

Langkah selanjutnya adalah Melakukan Standarisasi data dengan menggunakan metode Z-score, yaitu mengubah setiap nilai sehingga memiliki rata-rata 0 dan standar deviasi 1. Gambar 5 adalah tampilan Hasil Standarisasi dengan menggunakan *Pemrograman Phyton*

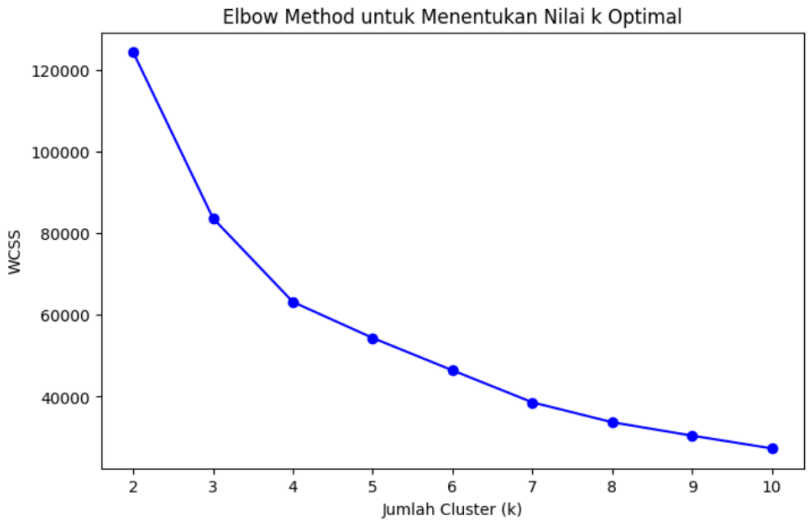
```
In [8]: from sklearn.preprocessing import StandardScaler

        # Standarisasi data numerik setelah encoding
        scaler = StandardScaler()
        df_scaled = scaler.fit_transform(df_encoded)
        df_scaled

Out[8]: array([[ 1.2994499 , -0.92816645, -0.949794  ,  0.94192601],
               [ 1.2994499 ,  1.07739296,  0.67935359, -0.22926188],
               [-1.15159982, -0.92816645,  0.67935359, -1.98604371],
               ...,
               [-1.44869675,  1.07739296,  0.67935359, -1.40044976],
               [-1.15159982,  1.07739296,  0.67935359,  0.94192601],
               [ 1.44799837,  1.07739296,  0.67935359, -0.22926188]])
```

Gambar 5. Hasil Standarisasi dengan menggunakan *Z-Score*

Menentukan jumlah cluster optimal dengan menggunakan metode Elbow. Metode ini menentukan jumlah cluster optimal dengan cara menghitung nilai *Within-Cluster Sum of Squares* (WCSS) untuk beberapa nilai k yaitu antara 2 hingga 10 cluster. Hasil dari metode tersebut didapatkan nilai K optimal adalah 4 dimana visualisasi hasil dapat dilihat pada Gambar 6.

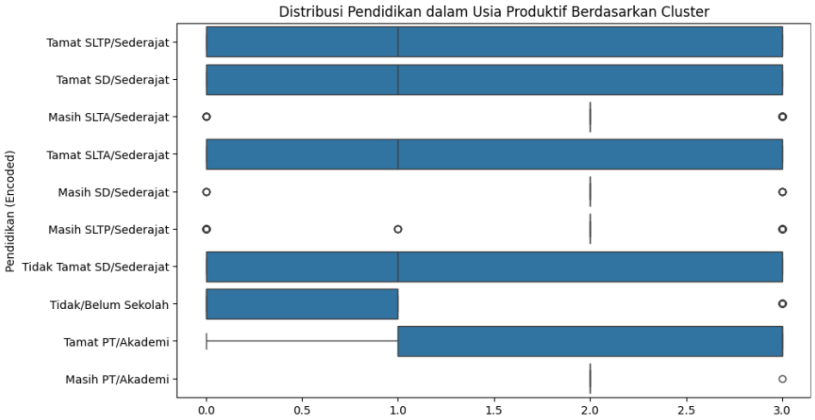


Gambar 6. Visualisasi grafik *Elbow Method*

Clustering dengan k-means menggunakan nilai k yang optimal yaitu 4 didapatkan hasil cluster pada keseluruhan data seperti yang terlihat pada Gambar 7 dan Gambar 8.

Cluster	Jumlah Anggota	Usia Rata-rata	Usia Min	Usia Max	Pekerjaan Dominan
0	16753	38.386259	15.0	64.0	Tidak/Belum Bekerja
1	6451	28.137653	15.0	64.0	Tidak/Belum Bekerja
2	7909	17.856998	15.0	57.0	Tidak/Belum Bekerja
3	15469	41.467839	16.0	64.0	Pekerja Lepas

Gambar 7. Hasil clustering dengan k-Means berdasarkan Usia dan Pekerjaan



Gambar 8. Grafik hubungan antara Pendidikan dengan usia Produktif

Hasil clustering berdasarkan hubungan antara Pendidikan dengan usia Produktif diperoleh bahwa Cluster 0 cenderung terdiri dari usia paruh baya hingga menjelang pensiun, tetapi masih dalam rentang usia produktif. Cluster 1 didominasi oleh usia muda dan masih banyak yang tidak/belum bekerja. Cluster 2 sebagian besar anggota tidak/belum bekerja, didominasi oleh remaja usia sekolah. Cluster 3 memiliki usia rata-rata paling tinggi, dengan mayoritas berada di usia paruh baya dan sebagian besar adalah pekerja lepas. Berdasarkan hasil 4 cluster ini didapatkan bahwa banyak individu dalam usia produktif belum bekerja (Cluster 0, 1, 2).

Distribusi Tingkat pendidikan dalam usia produktif berdasarkan cluster, didapatkan bahwa Mayoritas individu yang tidak/belum sekolah atau hanya tamat SD cenderung berada di Cluster 0 dan Cluster 1. Tingkat pendidikan menengah (SLTP dan SLTA) tersebar lebih merata di semua cluster. Individu dengan pendidikan tinggi (PT/Akademi) cenderung lebih terkonsentrasi di Cluster 3.

KESIMPULAN DAN SARAN

Berdasarkan hasil clustering K-Means yang dilakukan terhadap data penduduk dengan atribut Usia, Jenis Kelamin, Pekerjaan, dan Pendidikan. Menggunakan metode Elbow, nilai optimal K didapatkan adalah 4, yang menunjukkan pembagian kelompok yang paling sesuai berdasarkan distribusi data. Hasil clustering K-Means menunjukkan bahwa mayoritas penduduk usia produktif masih belum bekerja atau berada dalam pekerjaan informal. Terdapat empat cluster utama, di mana tiga di antaranya (Cluster 0, 1, dan 2) didominasi oleh individu yang tidak/belum bekerja, dengan usia rata-rata yang bervariasi antara 17 hingga 38 tahun, Cluster 3 memiliki dominasi pekerja lepas, denga rata-rata usia adalah 41,5 tahun.

Analisis lebih lanjut menunjukkan bahwa tingkat pendidikan sangat memengaruhi status pekerjaan. Mayoritas individu di Cluster 0 dan 1 hanya memiliki pendidikan SD atau SLTP, yang menjadi hambatan dalam mendapatkan pekerjaan formal. Sementara itu, Cluster 2 terdiri dari individu yang masih dalam tahap pendidikan, dan Cluster 3 lebih banyak dihuni oleh mereka dengan pendidikan lebih tinggi yang bekerja secara mandiri.

DAFTAR PUSTAKA

[1] Badan Pusat Statistik (BPS), "Statistik Penduduk Indonesia Tahun 2020."

[2] H. Xie *et al.*, "Improving K-means clustering with enhanced Firefly Algorithms," *Appl. Soft Comput. J.*, vol. 84, p. 105763, 2019, doi: 10.1016/j.asoc.2019.105763.

[3] C. D. L. Daniel T. Larose, *Discovering Knowledge in Data: An Introduction to Data Mining*. John Wiley & Sons, Inc., 2014. doi: DOI:10.1002/9781118874059.

[4] D. Triyansyah and D. Fitrianah, "Analisis Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Marketing," *J. Telekomun. dan Komput.*, vol. 8, no. 3, p. 163, 2018, doi: 10.22441/incomtech.v8i3.4174.

[5] F. Handayani, "Aplikasi Aplikasi Data Mining Menggunakan Algoritma K-Means Clustering untuk Mengelompokan Mahasiswa Berdasarkan Gaya Belajar," *J. Teknol. dan Inf.*, vol. 12, no. 1, pp. 46–63, 2022, doi: 10.34010/jati.v12i1.6733.

[6] P. Parjito and P. Permata, "Penerapan Data Mining Untuk Clustering Data Penduduk Miskin Menggunakan Algoritma Hard C-Means," *Data Manaj. dan Teknol. Inf.*, vol. 18, no. 1, pp. 64–69, 2019.

[7] R. Maulida, D. Haryadi, S. Ratu Counsela, and N. Rahmania Az Zahra, "Implementasi Algoritma K-Means dan Rapidminer untuk Clustering Data Kemiskinan Provinsi Banten," *J. Informatics Commun. Technol.*, vol. 6, no. 1, pp. 17–32, 2024.

[8] N. Nur Afidah, "Penerapan Metode Clustering dengan Algoritma K-means untuk Pengelompokkan Data Migrasi Penduduk Tiap Kecamatan di Kabupaten Rembang," *Prism. Pros. Semin. Nas. Mat.*, vol. 6, pp. 729–738, 2023, [Online]. Available: https://journal.unnes.ac.id/sju/index.php/prisma/

[9] Andi Diah Kuswanto, Azumardi Nabil Fadhila, Paulus Tri Setiawan, Muhammad Kevin Setiawan, and Dody Renal Syahputra, "Penerapan K-Means Clustering Untuk Menentukan Jumlah Pengangguran Berdasarkan Umur," *Repeater Publ. Tek. Inform. dan Jar.*, vol. 2, no. 3, pp. 135–146, 2024, doi: 10.62951/repeater.v2i3.116.

[10] R. R. Aria, "KLASTERISASI ANGKATAN KERJA DI INDONESIA BERDASARKAN USIA MENGGUNAKAN METODE ALGORITMA K-MEANS," *Inti Nusa Mandiri*, vol. 18, no. 2, pp. 157–165, 2024, doi: //doi.org/10.33480/inti.v18i2.5056.

[11] Darmansah, F. K. Putra, and T. N. Putra, "Penerapan Algoritma K-Means Clustering Pada Segmentasi Tenaga Kerja di Batam Dalam Sektor Manufaktur," *JSAI J. Sci. Appl. Informatics*, vol. 7, no. 3, pp. 522–530, 2024.

[12] Sugiyono, *Metode penelitian pendidikan : pendekatan kuantitatif, kualitatif, dan R&D*. ALFABETA, 2016.

[13] T. A. Lubis and M. Mulyati, "Prediksi Harga Saham Pfizer Inc (Fpe) Menggunakan Metode Regresi Linear Berganda Pfizer Inc (Fpe) Stock Price Prediction Using …," *J. Apl. Bisnis dan Komput.*, vol. 4, pp. 1–7, 2024, [Online]. Available: https://journal.unpak.ac.id/index.php/jubikom/article/view/9715

[14] N. A. Maori and E. Evanita, "Metode Elbow dalam Optimasi Jumlah Cluster pada K-Means Clustering," *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 14, no. 2, pp. 277–288, 2023, doi: 10.24176/simet.v14i2.9630.

[15] C. Yuan and H. Yang, "Research on K-Value Selection Method of K-Means Clustering Algorithm," *J*, vol. 2, no. 2, pp. 226–235, 2019, doi: 10.3390/j2020016.