



Implementasi Algoritma K-Means untuk Mengelompokkan Sosial Ekonomi Masyarakat Desa Cikampek Utara Berdasarkan Data Demografi Masyarakat

Dimas Fadilah¹, Baenil Huda², Elfina Novalia³, Tukino⁴

¹Program Studi Sistem Informasi, Universitas Buana Perjuangan Karawang, Indonesia, si22.dimasfadilah@mhs.ubpkarawang.ac.id

²Program Studi Sistem Informasi, Universitas Buana Perjuangan Karawang, Indonesia, baenil88@ubpkarawang.ac.id

³Program Studi Sistem Informasi, Universitas Buana Perjuangan Karawang, Indonesia, elfinanovalia@ubpkarawang.ac.id

⁴Program Studi Sistem Informasi, Universitas Buana Perjuangan Karawang, Indonesia, Tukino@ubpkarawang.ac.id

Corresponding Author: si22.dimasfadilah@mhs.ubpkarawang.ac.id¹

Abstract: *This study aims to classify the socio-economic conditions of the community in Cikampek Utara Village based on demographic data using the K-Means algorithm. The dataset consists of 14,892 residents and includes variables such as age, gender, marital status, and occupation. The research adopts a descriptive quantitative approach with several data processing stages, including preprocessing, transformation of categorical data into numerical form, normalization using Min-Max Scaling, and clustering based on Euclidean distance. The optimal number of clusters was determined using the Elbow Method and Silhouette Score, with the final selection of three clusters (K=3) based on a balance between model performance and interpretability. The results indicate that the population can be grouped into three main clusters: a transitional age group dominated by students and unemployed individuals, a domestic group contributing to household stability, and a productive workforce group acting as the main driver of the village economy. These findings demonstrate that the K-Means algorithm is effective in identifying socio-economic patterns and can support data-driven decision-making for targeted village development policies.*

Keywords: K-Means, clustering, data mining, demographic data, socio-economic, village development

Abstrak: Penelitian ini bertujuan untuk mengelompokkan kondisi sosial ekonomi masyarakat di Desa Cikampek Utara dengan memanfaatkan data demografi melalui penerapan algoritma K-Means. Data yang dianalisis mencakup 14.892 penduduk dengan variabel seperti usia, jenis kelamin, status pernikahan, dan pekerjaan. Metode yang digunakan adalah pendekatan kuantitatif deskriptif yang melibatkan beberapa tahapan pengolahan data, yaitu pra-pemrosesan, konversi data kategorikal menjadi numerik, normalisasi menggunakan metode Min-Max Scaling, serta proses pengelompokan berdasarkan jarak Euclidean.

Penentuan jumlah cluster terbaik dilakukan menggunakan metode Elbow dan Silhouette Score, yang menghasilkan tiga kelompok utama ($K=3$) dengan mempertimbangkan keseimbangan antara performa model dan kemudahan interpretasi. Hasil analisis menunjukkan bahwa masyarakat dapat diklasifikasikan ke dalam tiga kelompok, yaitu kelompok usia transisi yang didominasi oleh pelajar dan belum bekerja, kelompok domestik yang berperan dalam stabilitas rumah tangga, serta kelompok tenaga kerja produktif yang menjadi penggerak utama ekonomi desa. Temuan ini menunjukkan bahwa algoritma K-Means efektif dalam mengidentifikasi pola sosial ekonomi dan dapat digunakan sebagai dasar pengambilan keputusan berbasis data dalam perencanaan pembangunan desa.

Kata Kunci: K-Means, clustering, data mining, demografi, sosial ekonomi, pembangunan desa

PENDAHULUAN

Pembangunan di tingkat desa tidak hanya ditentukan oleh ketersediaan sumber daya fisik, tetapi juga dipengaruhi oleh kemampuan dalam memahami kondisi sosial ekonomi masyarakat. Informasi demografi seperti usia, jenis kelamin, pekerjaan, dan status pernikahan dapat memberikan gambaran mengenai karakteristik serta kebutuhan masyarakat. (Rafika et al., 2025) Namun, dalam praktiknya, data tersebut sering kali hanya dimanfaatkan sebagai arsip administratif tanpa diolah lebih lanjut untuk mendukung pengambilan keputusan. Hal ini menyebabkan potensi informasi yang terkandung di dalamnya belum dimanfaatkan secara maksimal. Kondisi serupa juga terjadi di Desa Cikampek Utara, di mana pendataan penduduk telah dilakukan secara rutin, tetapi belum diolah secara sistematis untuk mengidentifikasi pola sosial ekonomi masyarakat. (Muningsih et al., 2021)

Akibatnya, terdapat kesenjangan antara kondisi aktual dengan kondisi yang diharapkan. Saat ini, perangkat desa hanya menggunakan data penduduk untuk keperluan administrasi seperti pengurusan surat keterangan atau pelaporan jumlah penduduk. Apabila data tersebut diolah lebih lanjut, pemerintah desa dapat memperoleh informasi yang bermanfaat, misalnya kelompok masyarakat yang termasuk usia produktif tetapi belum bekerja, sebaran pekerjaan dominan, atau kelompok masyarakat rentan yang memerlukan dukungan ekonomi. (Humaira & Rasyidah, 2020) Informasi tersebut dapat menjadi dasar dalam penyusunan program pembangunan yang lebih tepat sasaran, seperti pelatihan keahlian, pemberdayaan UMKM, maupun perencanaan bantuan sosial. Oleh karena itu, diperlukan suatu metode yang mampu mengelompokkan data penduduk secara otomatis berdasarkan kemiripan karakteristiknya. (B. Huda et al., 2023)

Sejumlah penelitian terdahulu menunjukkan bahwa teknik data mining, khususnya algoritma K-Means Clustering, efektif digunakan untuk mengelompokkan data dengan karakteristik yang beragam. (Djaka Permana et al., 2023) Beberapa penelitian telah menerapkan *K-Means* untuk mengelompokkan data penduduk berdasarkan usia, tingkat pendidikan, jenis pekerjaan, atau kondisi ekonomi, dan hasilnya mampu membantu pengambilan kebijakan memahami pola yang sebelumnya tidak terlihat secara manual. *K-Means* bekerja dengan membagi data ke dalam beberapa kelompok (cluster) berdasarkan kedekatan nilai antar data, sehingga penduduk dengan karakteristik yang serupa akan berada dalam satu kelompok yang sama. (Trifena et al., 2023) Metode ini dianggap tepat karena prosesnya sederhana, waktu eksekusi cepat, dan dapat diterapkan pada data demografi yang tersedia di desa.

METODE

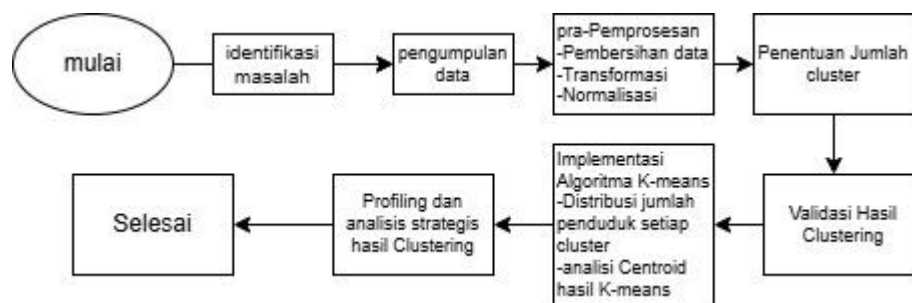
Objek penelitian ini adalah data demografi penduduk Desa Cikampek Utara yang berjumlah 14.892 jiwa. Data tersebut berasal dari dokumen resmi desa yang diperoleh

melalui izin penelitian dan berisi informasi dasar penduduk, seperti usia, jenis kelamin, pekerjaan, status pernikahan, dan jumlah anggota keluarga. Dataset ini digunakan untuk menganalisis karakteristik sosial ekonomi masyarakat berdasarkan variabel-variabel demografi yang tersedia.

Penelitian ini menggunakan pendekatan kuantitatif deskriptif karena bertujuan untuk mengelompokkan kondisi sosial ekonomi masyarakat berdasarkan data demografi menggunakan *algoritma K-Means*. (M. Huda & Hartati, 2022) Pendekatan kuantitatif digunakan karena seluruh proses analisis dilakukan melalui perhitungan matematis dan komputasional terhadap data numerik. (Wijaya et al., 2021)

Metode data mining dipilih karena mampu menemukan pola dan struktur tersembunyi dalam data kependudukan yang tidak dapat diidentifikasi melalui pengolahan data sederhana. Proses pengelompokan dilakukan menggunakan algoritma *K-Means* yang bekerja berdasarkan perhitungan jarak Euclidean dan proses iteratif terhadap centroid. (Rachmad, 2020)

Dengan demikian, jenis penelitian ini menggambarkan lingkup dan batasan yang jelas, yaitu memfokuskan analisis pada pengelompokan sosial ekonomi berdasarkan data demografi asli yang diperoleh dari pemerintah desa. (Wahono, 2023) Hasil penelitian diharapkan dapat memberikan gambaran kelompok masyarakat yang lebih akurat dan bermanfaat untuk perencanaan kebijakan di tingkat desa. (Widyanti et al., 2023)



Gambar 1 Penelitian Terkait
Sumber : (Ramdani et al., 2025)

Berdasarkan tahapan penelitian yang telah disajikan, setiap tahap memiliki peran yang saling berkaitan dalam mendukung proses penelitian. Penjelasan singkat masing-masing tahapan adalah sebagai berikut:

1. Identifikasi Masalah
Tahap ini dilakukan untuk menentukan permasalahan utama penelitian, yaitu perlunya pengelompokan kondisi sosial ekonomi masyarakat Desa Cikampek Utara berbasis data.
2. Pengumpulan Data
Data yang digunakan berupa data demografi penduduk yang diperoleh dari dokumen resmi pemerintah desa sebagai bahan analisis penelitian.
3. Pra-pemrosesan Data
Data mentah diproses melalui pembersihan, transformasi, dan normalisasi agar siap digunakan dalam proses komputasi algoritma *K-Means*. (Rofiqo et al., 2018)
4. Penentuan Jumlah Cluster
Tahap ini dilakukan untuk menentukan jumlah cluster yang digunakan dalam proses clustering agar hasil pengelompokan lebih optimal. (Muhamad Subur et al., 2024)
5. Validasi Hasil
Validasi hasil dilakukan menggunakan metode *Silhouette Coefficient* untuk menilai kualitas hasil pengelompokan yang diperoleh. (Listiani et al., 2019)
6. Implementasi Algoritma K-Means

Data yang telah diproses dianalisis menggunakan algoritma K-means untuk membentuk kelompok masyarakat berdasarkan kemiripan karakteristiknya. (Hariyanto et al., 2020)

7. Profiling dan Analisis Strategis Hasil Clustering

Profiling dan analisis strategis hasil clustering adalah proses mendeskripsikan karakteristik tiap kelompok yang terbentuk sekaligus menafsirkan maknanya untuk menentukan langkah atau kebijakan yang tepat. Profiling berfokus pada identifikasi ciri utama setiap cluster, seperti usia, status pernikahan, dan tingkat partisipasi ekonomi. Sementara itu, analisis strategis bertujuan merumuskan tindakan yang sesuai berdasarkan potensi dan kebutuhan masing-masing kelompok. (Tukino et al., 2023)

HASIL DAN PEMBAHASAN

Deskripsi Dataset Penelitian

Data yang digunakan dalam penelitian ini merupakan data demografi penduduk Desa Cikampek Utara yang diperoleh dari data administrasi desa. Dataset terdiri dari beberapa atribut yang relevan dengan kondisi sosial ekonomi masyarakat, yaitu usia, jenis kelamin, status pernikahan, dan pekerjaan. Data ini masih bersifat mentah dan mengandung atribut kategorik sehingga belum dapat langsung digunakan dalam proses clustering.

	nama_lengkap	tempat_lahir	usia	status_pernikahan	jenis_kelamin	alamat	pekerjaan
0	BASUKI ARACHMAN	MADIUN	71	S	L	KP.CISELANG	KARYAWAN SWASTA
1	RUTSAHANAYA MANURUNG	PANTAI PANJALAHAN	31	S	P	PERUM PERMATA REGENSI JL. ZAMBRUT 6 NO. 36	PETANI/PEKEBUN
2	AFDI TUA ANTON SINAMBELA	MEDAN	31	S	L	PERUM PERMATA REGENSI JL. ZAMBRUT 6 NO. 36	KARYAWAN SWASTA
3	RITA LIMBONG	LUMBAN SORMIN	32	B	P	PERUM REGENCY BLOK A20 NO.45 JL.MIRAH SIAM RAYA	WIRASWASTA
4	MEIRISKA BR GIRSANG	BANDUNG	21	B	P	JL PERMATA KASIH BLOK O.11 NO.20 REGENCY	PELAJAR/MAHASISWA

Gambar 2. Dataset Mentah

Hasil Pra-Pemrosesan Data

Tahap pra-pemrosesan data dilakukan untuk menyesuaikan data agar dapat diproses oleh algoritma K-Means. Pada tahap ini, atribut kategorik seperti jenis kelamin, status pernikahan, dan pekerjaan ditransformasikan ke dalam bentuk numerik menggunakan teknik label encoding. (Mutiah et al., 2024) Selanjutnya, seluruh atribut dinormalisasi menggunakan metode Min-Max Scaling agar memiliki rentang nilai yang seragam.

```
... array([[0.81538462, 0.      , 1.      , 0.32432432],
          [0.2      , 1.      , 1.      , 0.72972973],
          [0.2      , 0.      , 1.      , 0.32432432],
          [0.21538462, 1.      , 0.      , 1.      ],
          [0.04615385, 1.      , 0.      , 0.48648649]])
```

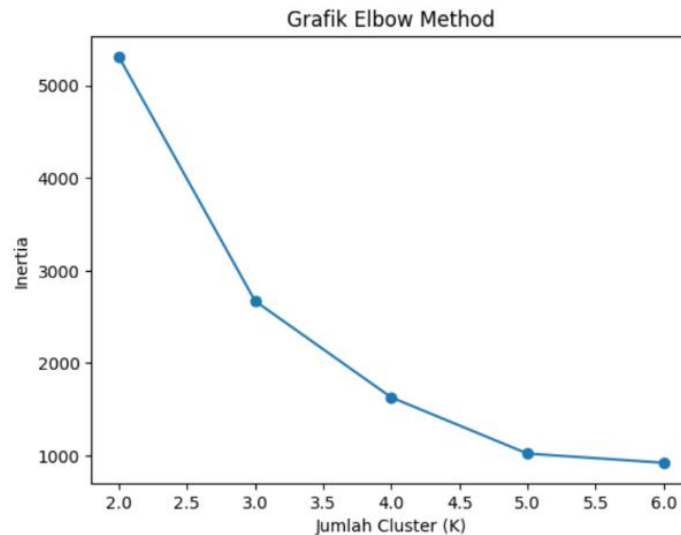
Gambar 3. Hasil Pra-pemrosesan Data

Proses pra-pemrosesan ini bertujuan untuk menghindari dominasi atribut tertentu dalam perhitungan jarak dan memastikan bahwa setiap atribut memiliki kontribusi yang seimbang dalam proses clustering. Seluruh proses pra-pemrosesan dilakukan secara otomatis menggunakan bahasa pemrograman Python pada platform Google Colab.

Penentuan Jumlah Cluster Optimal

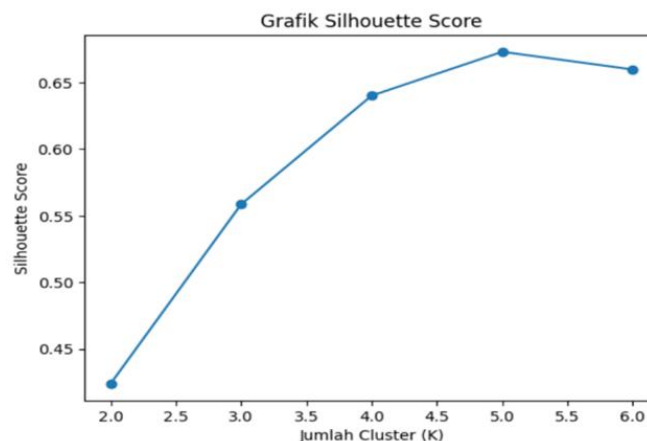
Penentuan jumlah cluster optimal merupakan tahap penting dalam proses clustering. Pada penelitian ini, penentuan jumlah cluster dilakukan menggunakan dua metode evaluasi, yaitu Elbow Method dan Silhouette Score.

Berdasarkan grafik Elbow Method yang ditampilkan pada Gambar 4.3, terlihat adanya penurunan nilai inerti yang signifikan hingga jumlah cluster $K=3$, kemudian penurunannya mulai melandai. Pola ini menunjukkan titik siku (elbow) pada $K=3$, yang mengindikasikan jumlah cluster optimal.



Gambar 4. Grafik Elbow Method

Validasi tambahan dilakukan menggunakan Silhouette Score untuk mengukur kualitas pemisahan antar cluster. Nilai silhouette meningkat seiring bertambahnya jumlah cluster dan mencapai nilai maksimum pada $K=5$.



Gambar 5. Grafik Silhouette Score

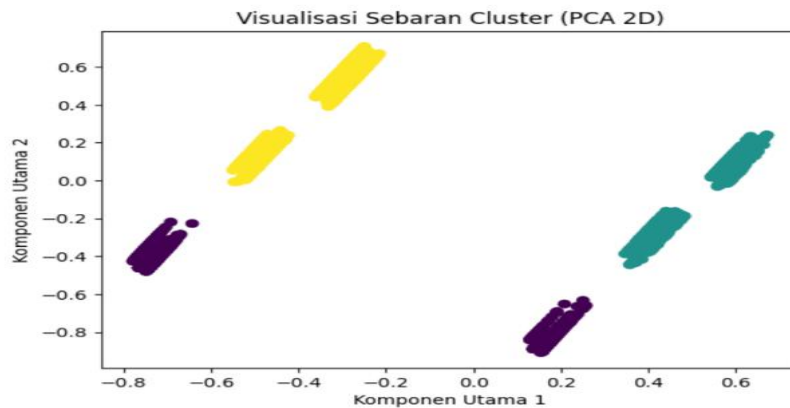
Namun demikian, dalam konteks penelitian ini, pemilihan jumlah cluster tidak hanya mempertimbangkan nilai evaluasi numerik semata, tetapi juga aspek interpretabilitas dan relevansi terhadap analisis sosial ekonomi masyarakat desa. Dengan mempertimbangkan keseimbangan antara kualitas model dan kemudahan interpretasi hasil segmentasi, jumlah cluster ditetapkan sebanyak tiga ($K=3$). (Gymnastiar & Bahtiar, 2024) Pemilihan ini dinilai cukup representatif untuk menggambarkan struktur sosial ekonomi masyarakat Desa Cikampek Utara secara jelas, tanpa menghasilkan segmentasi yang terlalu kompleks sehingga

menyulitkan perumusan implikasi kebijakan.

Implementasi K-Means

Setelah jumlah cluster optimal ditentukan, proses clustering dilakukan menggunakan algoritma K-Means dengan jumlah cluster K=3. Untuk memudahkan interpretasi hasil clustering, dilakukan visualisasi menggunakan metode Principal Component Analysis (PCA) dua dimensi. (Rahayu et al., 2018)

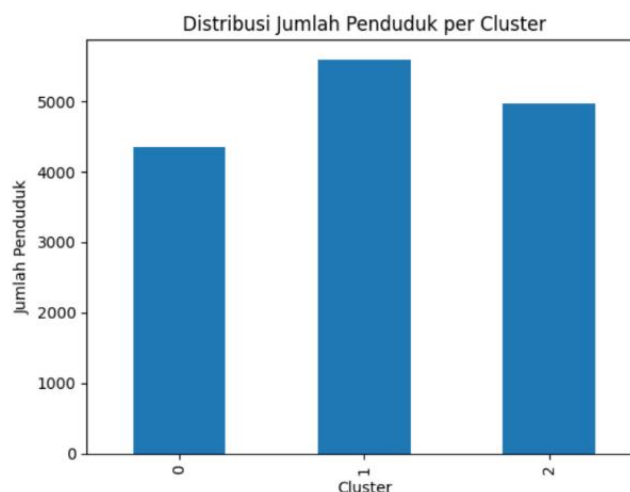
Hasil visualisasi PCA pada Gambar 4.5 menunjukkan bahwa data penduduk terbagi ke dalam tiga cluster yang terpisah dengan cukup jelas. Hal ini menandakan bahwa algoritma K-Means mampu mengelompokkan data berdasarkan kemiripan karakteristik demografis secara efektif.



Gambar 6. Visualisasi Sebaran Cluster (PCA 2D)

Distribusi Jumlah Penduduk Pada Setiap Cluster

Distribusi jumlah penduduk pada masing masing cluster ditampilkan pada gambar, grafik tersebut menunjukkan bahwa jumlah anggota pada setiap cluster relatif seimbang. Hal ini menunjukkan bahwa hasil clustering tidak didominasi oleh satu cluster tertentu dan mampu merepresentasikan kondisi masyarakat secara proporsional.



Gambar 7. Distribusi jumlah penduduk per cluster

Analisis Centroid Hasil Clustering K-Means

Table 1. Nilai Centroid Hasil Clustering

Cluster	Usia	Jenis kelamin	Status Pernikahan	Pekerjaan
0	0.1208	0.3972	0.0000	0.4107
1	0.4277	1.0000	0.9265	0.3921

2	0.4552	0.0000	0.9679	0.4518
---	--------	--------	--------	--------

Nilai centroid pada Tabel merepresentasikan nilai rata-rata setiap variabel dalam masing-masing cluster setelah melalui proses normalisasi menggunakan metode MinMaxScaler dengan rentang 0–1. Dalam konteks ini, nilai yang mendekati 0 menunjukkan kecenderungan rendah terhadap suatu variabel, sedangkan nilai yang mendekati 1 menunjukkan kecenderungan tinggi. Oleh karena itu, interpretasi centroid tidak dibaca sebagai angka absolut, melainkan sebagai pola relatif antar cluster.

Berdasarkan variabel usia, Cluster 0 memiliki nilai paling rendah (0.120827) yang menunjukkan dominasi kelompok usia lebih muda, sedangkan Cluster 1 (0.427763) dan Cluster 2 (0.455271) berada pada rentang usia produktif yang lebih tinggi. Pada variabel jenis kelamin, Cluster 1 menunjukkan dominasi satu kategori secara penuh (nilai 1), sementara Cluster 2 didominasi kategori kebalikannya (nilai mendekati 0), dan Cluster 0 memiliki komposisi yang lebih seimbang. Variabel status pernikahan menjadi pembeda paling signifikan, di mana Cluster 0 didominasi individu belum menikah (≈ 0), sedangkan Cluster 1 (0.926552) dan Cluster 2 (0.967950) hampir seluruhnya terdiri dari individu yang telah menikah. Pada variabel pekerjaan, Cluster 2 memiliki nilai tertinggi (0.451811) yang mengindikasikan tingkat aktivitas ekonomi lebih tinggi dibandingkan dengan Cluster 0 (0.410759) dan Cluster 1 (0.392112).

Secara keseluruhan, pola nilai centroid menunjukkan bahwa algoritma K-Means berhasil memisahkan masyarakat berdasarkan fase kehidupan, struktur gender, status pernikahan, dan tingkat aktivitas ekonomi secara konsisten dan signifikan.

Profiling dan Analisis Strategis Hasil Clustering

Hasil Pengelompokan masyarakat Desa Cikampek Utara menggunakan algoritma K-Means menunjukkan adanya diferensiasi struktur sosial ekonomi yang jelas, terukur, dan sistematis. Pembentukan tiga cluster tidak hanya merepresentasikan pembagian numerik semata, tetapi menggambarkan segmentasi demografis berdasarkan fase kehidupan, struktur rumah tangga, serta tingkat partisipasi ekonomi masyarakat. Temuan ini konsisten dengan nilai centroid yang menunjukkan bahwa variable usia, status, pernikahan, dan jenis kelamin menjadi faktor pembeda utama dalam pembentukan cluster. Dengan demikian, struktur kelompok yang terbentuk memiliki dasar matematis sekaligus relevansi sosial yang kuat.

Table 2. Distribusi Jenis Pekerjaan Pada Setiap Cluster

Cluster	Pekerjaan	Jumlah
Cluster 0	Pelajar/mahasiswa	2860
	Belum/Tidak bekerja	679
	Karyawan Swasta	322
Cluster 1	Mengurus Rumah Tangga	4588
	Karyawan Swasta	469
	Wiraswasta	206
Cluster 2	Buruh Hrian Lepas	166
	Karyawan Swasta	1483
	Wiraswasta	1412

1. Cluster 0 Generasi Transisi (Early Economic Stage)

Cluster 0 didominasi oleh kategori pelajar/mahasiswa (2860 orang) serta individu belum/tidak bekerja (679 orang). Berdasarkan nilai centroid, cluster ini memiliki rata-rata usia paling rendah, mayoritas berstatus belum menikah, serta komposisi gender yang relatif seimbang.

Secara sosial ekonomi, kelompok ini berada pada fase transisi dari ketergantungan ekonomi menuju kemandirian. Dominasi pelajar menunjukkan bahwa desa memiliki cadangan sumber daya manusia jangka panjang yang signifikan. Namun, tingginya jumlah

individu yang belum bekerja juga mengindikasikan potensi risiko pengangguran usia muda apabila tidak didukung oleh kebijakan peningkatan kompetensi.

Secara struktural, cluster ini merepresentasikan fase awal dalam siklus ekonomi desa dan berperan sebagai generasi penerus pembangunan. Oleh karena itu, diperlukan intervensi kebijakan yang berorientasi pada penguatan kapasitas dan kesiapan kerja kelompok ini. Program pelatihan vokasi yang disesuaikan dengan kebutuhan pasar kerja, penguatan keterampilan digital, serta pembinaan kewirausahaan bagi pemuda menjadi langkah strategis untuk meningkatkan daya saing mereka. Selain itu, pengembangan inkubator UMKM pemula dan pendampingan karier berbasis komunitas dapat membantu mempercepat proses transisi dari fase pendidikan menuju kemandirian ekonomi. Dengan pendekatan tersebut, cluster ini dapat dipandang sebagai investasi sosial jangka panjang yang memiliki potensi signifikan dalam mendukung keberlanjutan pembangunan ekonomi desa.

2. Cluster 1 Kelompok Stabil Domestik (Domestic Economic Support)

Cluster 1 sangat didominasi oleh kategori mengurus rumah tangga (4588 orang). Secara demografis, kelompok ini didominasi perempuan, berstatus menikah, dan berada pada usia produktif. Secara sosial ekonomi, cluster ini berfungsi sebagai stabilisator rumah tangga. Meskipun tidak seluruhnya terlibat dalam sektor kerja formal, kelompok ini memiliki potensi ekonomi laten yang cukup besar. Dominasi aktivitas domestik membuka peluang pengembangan ekonomi berbasis rumah tangga seperti industri rumahan, kuliner, maupun kerajinan.

Dalam struktur sosial desa, cluster ini berperan penting dalam menjaga keberlanjutan sosial dan stabilitas keluarga yang menjadi fondasi utama pembangunan. Meskipun aktivitas ekonominya banyak berlangsung di ranah domestik, kelompok ini memiliki potensi signifikan untuk dikembangkan menjadi kekuatan ekonomi mikro yang produktif. Oleh karena itu, diperlukan kebijakan yang mendukung pemberdayaan ekonomi perempuan melalui pelatihan UMKM skala mikro, penguatan kapasitas produksi, serta pemanfaatan digital marketing berbasis rumah tangga untuk memperluas akses pasar. Selain itu, penyediaan akses permodalan usaha mikro menjadi faktor penting dalam mendorong kemandirian ekonomi kelompok ini. Dengan intervensi kebijakan yang tepat sasaran, cluster ini dapat berkembang menjadi motor penggerak ekonomi mikro desa yang berkontribusi terhadap peningkatan kesejahteraan masyarakat secara berkelanjutan.

3. Cluster 2 Tulang Punggung Ekonomi Desa (Core Economic Workforce)

Cluster 2 didominasi oleh buruh harian lepas (1666 orang), karyawan swasta (1482 orang), dan wiraswasta (1412 orang). Berdasarkan karakteristik demografis, kelompok ini berada pada usia matang dan produktif, mayoritas telah menikah, serta didominasi laki-laki. Secara sosial ekonomi, cluster ini merupakan kontributor utama dalam perputaran ekonomi desa. Tingginya jumlah pekerja sektor informal dan semi-formal menunjukkan bahwa pendapatan kelompok ini berpotensi fluktuatif serta rentan terhadap dinamika ekonomi eksternal. Namun demikian, kelompok ini menjadi penggerak utama aktivitas ekonomi, baik melalui tenaga kerja maupun usaha mandiri.

Dalam struktur sosial ekonomi desa, cluster ini berperan sebagai penopang utama ekonomi rumah tangga sekaligus penggerak aktivitas ekonomi lokal. Kontribusinya terhadap perputaran ekonomi desa menjadikan kelompok ini sebagai elemen strategis dalam menjaga stabilitas dan keberlanjutan pembangunan. Oleh karena itu, diperlukan kebijakan yang mampu memperkuat kapasitas ekonomi kelompok ini, seperti peningkatan akses permodalan bagi pelaku UMKM, penyediaan program jaminan sosial bagi tenaga kerja sektor informal, serta pelatihan peningkatan keterampilan teknis guna meningkatkan produktivitas dan daya saing. Selain itu, program stabilisasi ekonomi keluarga juga menjadi penting untuk menjaga ketahanan ekonomi rumah tangga. Dengan dukungan

kebijakan yang tepat, cluster ini dapat dikategorikan sebagai pilar utama keberlanjutan ekonomi Desa Cikampek Utara.

KESIMPULAN

Berdasarkan hasil analisis dan pembahasan yang telah dilakukan terhadap data demografi masyarakat Desa Cikampek Utara menggunakan algoritma K-Means, dapat disimpulkan bahwa metode clustering mampu mengelompokkan masyarakat ke dalam tiga segmen sosial ekonomi yang berbeda secara jelas dan terukur. Proses penentuan jumlah cluster dilakukan melalui pendekatan Elbow Method dan Silhouette Score, yang menunjukkan bahwa meskipun nilai evaluasi numerik tertinggi diperoleh pada jumlah cluster yang lebih besar, pemilihan tiga cluster (K=3) dinilai paling representatif karena memberikan keseimbangan antara kualitas pemisahan data dan kemudahan interpretasi dalam konteks sosial ekonomi desa.

Hasil pengelompokan menunjukkan adanya struktur sosial ekonomi yang terbagi ke dalam tiga karakteristik utama, yaitu kelompok usia transisi yang didominasi pelajar dan individu belum bekerja, kelompok domestik stabil yang didominasi pengelola rumah tangga dengan status pernikahan tinggi, serta kelompok tenaga kerja produktif yang menjadi penggerak utama aktivitas ekonomi desa. Analisis nilai centroid memperlihatkan bahwa variabel usia, status pernikahan, dan jenis kelamin menjadi faktor pembeda dominan dalam pembentukan cluster, sedangkan variabel pekerjaan memperkuat segmentasi ekonomi antar kelompok yang terbentuk.

Secara keseluruhan, penelitian ini membuktikan bahwa penerapan algoritma K-Means pada data demografi tidak hanya menghasilkan pengelompokan berbasis komputasi, tetapi juga mampu merepresentasikan pola sosial ekonomi masyarakat secara kontekstual dan aplikatif. Dengan demikian, hasil clustering dapat dijadikan sebagai dasar pendukung dalam perencanaan program pembangunan desa yang lebih terarah dan berbasis data.

REFERENSI

- Djaka Permana, M., Lia Hananto, A., Novalia, E., Huda, B., & Paryono, T. (2023). Klasterisasi Data Jamaah Umrah pada Tanurmutmainah Tour Menggunakan Algoritma K-Means. *Jurnal KomtekInfo*, 10, 15–20. <https://doi.org/10.35134/komtekinfo.v10i1.332>
- Gymnastiar, S., & Bahtiar, A. (2024). Penerapan Algoritma K-Means Clustering Untuk Mengelompokkan Data Kejadian Kekeringan Di Kabupaten Cirebon. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(2), 2325–2331. <https://doi.org/10.36040/jati.v8i2.8948>
- Hariyanto, M., Kholiq, M., Yani, A., & Narti. (2020). Penerapan Algoritma K-means untuk klasifikasi klasterisasi penduduk miskin di provinsi Banten. *Inti Nusa Mandiri*, 14(2), 133–138. <https://doi.org/https://doi.org/10.33480/inti.v18i1.4399> VOL.
- Huda, B., Manongga, D., Sedyono, E., Yulianto, S., Fauzi, A., Hananto, A. L., Tukino, & Tarmuji. (2023). Implementation of UI/UX the Design Thinking Approach Method In Inventory Information System. *E3S Web of Conferences*, 448, 02005. <https://doi.org/10.1051/e3sconf/202344802005>
- Huda, M., & Hartati, N. (2022). Implementasi Strategi Terhadap Supply Chain Management, Keunggulan Bersaing Dan Kinerja Perusahaan. *Jurnal Soshum Insentif*, 5(1), 28–35. <https://doi.org/10.36787/jsi.v5i1.646>
- Humaira, H., & Rasyidah, R. (2020). Determining The Appropriate Cluster Number Using Elbow Method for K-Means Algorithm. *Journal of Telkom University*, 2(January 2020). <https://doi.org/10.4108/eai.24-1-2018.2292388>
- Listiani, L., Agustin, Y. H., & Ramdhani, M. Z. (2019). Implementasi algoritma k-means cluster untuk rekomendasi pekerjaan berdasarkan pengelompokan data penduduk. *Seminar Nasional Sistem Informasi Dan Teknik Informatika*, 761–769.
- Muhamad Subur, Martanto, & Umi Hayati. (2024). Clustering Penduduk Miskin Untuk Penerima Bantuan Sosial Menggunakan Algoritma K-Means (Studi Kasus: Desa

- Greged). *JATI: Jurnal Mahasiswa Teknik Informatika*, 8(1), 789–795.
- Muningsih, E., Maryani, I., & Handayani, V. R. (2021). Penerapan Metode K-Means dan Optimasi Jumlah Cluster dengan Index Davies Bouldin untuk Clustering Propinsi Berdasarkan Potensi Desa. *Jurnal Sains Dan Manajemen*, 9(1), 96. www.bps.go.id
- Mutiah, S., Hasnataeni, Y., Fitrianto, A., & Jumansyah, L. M. R. D. (2024). Perbandingan Metode Klastering K-Means dan DBSCAN dalam Identifikasi Kelompok Rumah Tangga Berdasarkan Fasilitas Sosial Ekonomi di Jawa Barat Dalam era digital saat ini , jumlah data yang tersedia dari berbagai bidang , termasuk sosial dan ekonomi , terus. *Teorema: Teori Dan Riset Matematika*, 09(September), 247–260.
- Rachmad, D. S. (2020). Klasifikasi pada Tempat Tinggal Menurut Provinsi dan Jenis Kepemilikan Berdasarkan Algoritma K-Means. *STRING (Satuan Tulisan Riset Dan Inovasi Teknologi)*, 4(3), 247. <https://doi.org/10.30998/string.v4i3.5932>
- Rafika, A., Putri, H., & Wakhidah, N. (2025). Optimization of K-Means Clustering with Elbow Method for Identification of TB Prone in Central Java. *Journal of Computer Science, Information Technology and Telecommunication Engineering*, 6(1), 730–736. <https://doi.org/10.30596/jcositte.v6i1.21669>
- Rahayu, P., Sudipa, I. G. I., Suryani, Surachman, A., Ridwan, A., Darmawiguna, I. G. M., Sutoyo, M., Slamet, I., Harlina, S., & May Sanjaya, I. M. (2018). *Buku Ajar Data Mining* (Efitra (ed.); Cetakan Pe, Vol. 1, Issue January 2024). PT. Sonpedia Publishing Indonesia. www.buku.sonpedia.com
- Ramdani, R., Suarna, N., Ali, I., & Efendi, D. I. (2025). Penerapan Algoritma K-Means Dalam Analisis Data Kependudukan Untuk Optimalisasi Pengelompokan Di Desa Pasawahan. *Jurnal Informatika Dan Teknik Elektro Terapan*, 13(1), 439–445. <https://doi.org/10.23960/jitet.v13i1.5639>
- Rofiqo, N., Windarto, A. P., & Hartama, D. (2018). Penerapan Clustering Pada Penduduk Yang Mempunyai Keluhan Kesehatan Dengan Datamining K-Means. *KOMIK (Konferensi Nasional Teknologi Informasi Dan Komputer)*, 2(1), 216–223. <https://doi.org/10.30865/komik.v2i1.929>
- Trifena, M., Voutama, A., & Ridha, A. A. (2023). Perancangan Ui/Ux Aplikasi Sistem Pendaftaran Rumah Sakit Saraswati Berbasis Mobile Dengan Metode Design Thinking. *Information Management for Educators and Professionals*, 7(2), 113–123.
- Tukino, Muhmaad Rizky Arfani, Agustia Hananto, B. H. (2023). Seleksi Penerimaan Bantuan Internet Gratis dengan Menggunakan Metode AHP. *Internal Information System*, 6(1), 1–9. <https://doi.org/10.32627>
- Wahono, R. S. (2023). Data Mining Data mining. In M. K. Dina ediana, S. Kom. (Ed.), *Mining of Massive Datasets* (Cetakan Pe, Vol. 2, Issue January 2013). PT GLOBAL EKSEKUTIF TEKNOLOGI. https://www.cambridge.org/core/product/identifier/CBO9781139058452A007/type/book_part
- Widyanti, T., Hilabi, S. S., Hananto, A., Tukino, & Novalia, E. (2023). Implementasi K-Means dan K-Nearest Neighbors pada Kategori Siswa Berprestasi. *Jurnal Informasi Dan Teknologi*, 5(1), 75–82. <https://doi.org/10.37034/jidt.v5i1.255>
- Wijaya, R. S. D., Adiwijaya, Andriyan B Suksmono, & Tati LR Mengko. (2021). Segmentasi Citra Kanker Serviks Menggunakan Markov Random Field dan Algoritma K-Means. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 5(1), 139–147. <https://doi.org/10.29207/resti.v5i1.2816>