



Department of Digital Business

**Journal of Artificial Intelligence and Digital Business (RIGGS)**

Homepage: <https://journal.ilmudata.co.id/index.php/RIGGS>

Vol. 4 No. 2 (2025) pp: 5050-5055

P-ISSN: 2963-9298, e-ISSN: 2963-914X

---

## Analisis Kluster Kekayaan Tokoh Terkaya Indonesia Tahun 2007 Menggunakan Algoritma K-Means

Rizky Ichwan Purnama<sup>1</sup>, Lutfan Hakim<sup>2</sup>, Aryansyah Al Giffari<sup>3</sup>, Faizal Andre Febrianto<sup>4</sup>, Farizi Ilham<sup>5</sup>

<sup>1</sup>Program Studi Teknik Informatika, Universitas Pamulang

<sup>1</sup>rizkyip1306@gmail.com, <sup>2</sup>lutfanhakim777@gmail.com, <sup>3</sup>aryansmail03@gmail.com, <sup>4</sup>faizalandrefebrianto@gmail.com,

<sup>5</sup>dosen02954@unpam.ac.id

### **Abstrak**

*Penelitian ini bertujuan untuk mengelompokkan individu dalam daftar 40 orang terkaya di Indonesia tahun 2007 versi Forbes berdasarkan jumlah kekayaan dan sektor usahanya menggunakan algoritma K-Means Clustering. Dengan memanfaatkan data numerik kekayaan dan menerapkan metode pra-pemrosesan yang sesuai, dilakukan analisis untuk menentukan jumlah kluster optimal menggunakan Elbow Method dan Silhouette Score. Hasil analisis menunjukkan bahwa pembagian ke dalam tiga kluster merupakan struktur yang paling representatif. Kluster pertama terdiri dari individu dengan kekayaan sangat tinggi yang memiliki diversifikasi usaha di berbagai sektor strategis. Kluster kedua mencakup pengusaha mapan dengan fokus sektor tunggal, sementara kluster ketiga mencerminkan kelompok pengusaha yang kekayaannya relatif lebih kecil namun memiliki potensi pertumbuhan. Studi ini menunjukkan bahwa algoritma K-Means efektif dalam mengidentifikasi pola distribusi kekayaan dan dapat menjadi alat bantu dalam analisis ekonomi dan bisnis berbasis data.*

*Kata kunci: K-Means, Klasterisasi, Orang Terkaya Indonesia, Elbow Method, Silhouette Score*

### **1. Latar Belakang**

Distribusi kekayaan dalam suatu bangsa mencerminkan kerumitan sistem ekonomi, kebijakan fiskal dan moneter, serta dinamika sosial-politik yang membentuk hubungan antara pelaku ekonomi. Fenomena ini tidak hanya terkait dengan pengumpulan kekayaan oleh individu atau kelompok tertentu, tetapi juga bersinggungan dengan isu-isu fundamental seperti keadilan sosial, peluang ekonomi, dan stabilitas negara. Ketimpangan dalam distribusi kekayaan telah lama menjadi perhatian dalam kajian ekonomi karena dampaknya yang luas terhadap struktur masyarakat, mulai dari pengaruh pada daya beli hingga stabilitas politik dan penerimaan masyarakat terhadap pemerintahan. Oleh karena itu, memahami bagaimana kekayaan didistribusikan serta alasan di balik ketimpangan tersebut adalah isu yang sangat penting.

Indonesia, sebagai negara berkembang dengan karakteristik geografis kepulauan dan populasi yang besar, menghadapi tantangan tersendiri dalam hal pemerataan pembangunan dan distribusi kekayaan. Sejak reformasi setelah 1998, pertumbuhan ekonomi Indonesia menunjukkan tren positif yang berkelanjutan, dengan berbagai sektor bisnis berkembang. Namun, pertumbuhan yang tinggi tidak selalu sejalan dengan pemerataan kesejahteraan. Dalam banyak kasus, pertumbuhan ekonomi seringkali lebih banyak dinikmati oleh sekelompok kecil elit ekonomi yang memiliki akses terhadap sumber daya, informasi, dan jaringan kekuasaan. Dalam konteks ini, penting untuk mengkaji lebih dalam mengenai masalah konsentrasi kekayaan.

Salah satu metode untuk memahami struktur kekayaan negara adalah dengan menganalisis data orang-orang terkaya, seperti yang dipublikasikan secara reguler oleh majalah Forbes. Data ini, meskipun tidak mencakup seluruh populasi, memberikan gambaran penting mengenai lapisan atas dalam piramida ekonomi. Pada tahun 2007, Indonesia berada dalam tahap pemulihan ekonomi yang cukup stabil setelah krisis finansial Asia tahun 1997-1998. Tahun tersebut menjadi momen penting dalam sejarah ekonomi Indonesia, menandai awal konsolidasi kekuatan ekonomi baru, terutama dari sektor-sektor yang sebelumnya kurang dominan.

Data Forbes yang memuat daftar 40 orang terkaya di Indonesia pada tahun 2007 menjadi sumber yang sangat berharga untuk dianalisis. Selain mencantumkan total nilai kekayaan masing-masing individu, data ini juga menyajikan informasi tentang sektor usaha yang menjadi sumber utama kekayaan. Informasi ini memiliki potensi

untuk analisis yang tidak hanya bersifat deskriptif, tetapi juga dapat diarahkan pada identifikasi pola-pola akumulasi kekayaan dan hubungan antara sektor tertentu dengan tingkat kekayaan yang berhasil dikumpulkan. Dari sini muncul berbagai pertanyaan yang mendalam: apakah individu yang bergerak dalam sektor sumber daya alam cenderung memiliki kekayaan lebih dibandingkan yang berada di sektor manufaktur atau jasa? Apakah terdapat pola tertentu dalam diversifikasi usaha yang dilakukan oleh para miliarder tersebut?

Untuk menjawab pertanyaan-pertanyaan ini, diperlukan pendekatan analisis data yang dapat mengungkap struktur dan pola yang tersembunyi dalam kumpulan data kekayaan tersebut. Salah satu pendekatan yang relevan dalam konteks ini adalah teknik penambangan data, khususnya metode klasterisasi. Klasterisasi adalah teknik analisis data yang bertujuan untuk mengelompokkan objek berdasarkan kesamaan karakteristik mereka. Teknik ini bersifat tidak terawasi, yang berarti tidak memerlukan label atau kategori sebelumnya, sehingga cocok digunakan ketika peneliti ingin menemukan pola-pola alami dalam data yang masih belum terdefinisi.

Di antara beragam algoritma klasterisasi yang ada, K-Means merupakan salah satu yang paling banyak digunakan karena kesederhanaannya dalam pengaplikasian serta kemampuannya untuk menghasilkan pengelompokan yang cukup tepat secara efisien. Algoritma ini melakukan partisi pada sekumpulan data menjadi beberapa klaster berdasarkan kedekatan jarak dengan titik pusat (centroid) dari setiap klaster. Namun, penerapan K-Means pada data kekayaan memerlukan langkah-langkah pra-pemrosesan yang teliti, terutama karena data terdiri dari kombinasi numerik (jumlah kekayaan) dan kategori (jenis usaha). Langkah normalisasi dan pengkodean adalah tahap awal yang penting agar algoritma dapat berfungsi dengan baik.

Selain sebagai kegiatan statistik, penerapan K-Means pada data orang-orang terkaya di Indonesia dapat memberikan wawasan baru tentang kondisi ekonomi nasional pada waktu itu. Klaster yang dihasilkan mungkin mengindikasikan bahwa kekayaan di Indonesia pada tahun 2007 lebih banyak terakumulasi di sektor-sektor tertentu, seperti tambang, agribisnis, atau perbankan. Alternatifnya, hasil klasterisasi juga bisa menunjukkan bahwa individu-individu terkaya di Indonesia memiliki ciri-ciri yang sangat bervariasi, baik dalam ukuran kekayaan maupun sektor usaha, yang bisa menjadi indikator tingginya diversifikasi ekonomi di kalangan elite ini.

Dengan pemetaan klaster yang terbentuk, peneliti dan pengambil kebijakan bisa mendapatkan wawasan lebih dalam mengenai struktur elite ekonomi di Indonesia. Apakah terdapat akumulasi kekayaan pada kelompok tertentu yang dapat mempengaruhi proses pengambilan kebijakan ekonomi? Apakah ada sektor yang mendominasi pengumpulan kekayaan sehingga perlu ada pengaturan lebih ketat untuk mencegah praktik monopoli atau kartel? Pertanyaan-pesannya ini semakin relevan ketika hasil klasterisasi bisa memberikan gambaran yang jelas tentang pola distribusi kekayaan dan arah pembangunan ekonomi.

Oleh karena itu, penelitian ini tidak hanya bertujuan mengimplementasikan algoritma K-Means pada data kekayaan individu terkaya di Indonesia pada tahun 2007, tetapi juga berupaya memberikan kontribusi dalam membangun pemahaman kritis tentang dinamika ekonomi serta ketimpangan struktural di Indonesia. Melalui pendekatan ini, diharapkan bisa didapatkan gambaran yang lebih komprehensif mengenai bagaimana kekayaan terdistribusi, sektor-sektor apa yang menjadi pendorong utama kekayaan, dan bagaimana hal ini berpengaruh terhadap tatanan sosial ekonomi yang lebih luas. Penelitian ini juga menunjukkan bahwa ilmu data, khususnya teknik klasterisasi, dapat menjadi alat yang sangat bermanfaat dalam menganalisis fenomena ekonomi serta merumuskan kebijakan berdasarkan data yang ada.

## 2. Metode Penelitian

Bagian metodologi ini menjelaskan secara sistematis langkah-langkah yang diambil dalam penelitian, mulai dari pengumpulan data hingga implementasi algoritma klasterisasi. Seluruh proses analisis dan komputasi dilakukan menggunakan bahasa pemrograman Python dengan memanfaatkan library khusus untuk data science dan machine learning. Ketepatan dan rincian dalam metodologi sangat penting untuk memastikan reproduktivitas dan validitas penelitian [3].

### 2.1. Sumber data

Data yang digunakan dalam penelitian ini bersumber dari daftar "40 Orang Terkaya di Indonesia Tahun 2007" yang dipublikasikan secara resmi oleh majalah Forbes. Forbes dikenal sebagai salah satu media terkemuka yang secara berkala merilis daftar kekayaan individu di berbagai negara, sehingga menjamin kredibilitas dan keandalan data. Informasi spesifik yang diekstrak dari daftar tersebut meliputi:

1. Nama Individu: Sebagai identifikasi unik untuk setiap entitas dalam dataset.
2. Jumlah Kekayaan: Disajikan dalam satuan miliar USD, merepresentasikan variabel numerik utama untuk klasterisasi.
3. Sumber Kekayaan: Memberikan gambaran umum mengenai asal-usul atau basis utama kekayaan individu.

4. Bidang Usaha Utama: Menggambarkan sektor ekonomi spesifik tempat individu tersebut aktif berinvestasi atau menjalankan bisnisnya. Informasi ini dapat bersifat kategorial dan memerlukan pra-pemrosesan lebih lanjut.

Data ini kemudian diimpor dan dikelola dalam Python, umumnya menggunakan *library* Pandas untuk struktur data tabular.

## 2.2. Pra pemrosesan data

Tahap pra-pemrosesan data krusial untuk memastikan data siap diolah oleh algoritma K-Means. Kualitas hasil klusterisasi sangat bergantung pada tahapan ini [4]. Seluruh proses pra-pemrosesan ini diimplementasikan menggunakan Python.

- Normalisasi Data Kekayaan: Variabel 'Jumlah Kekayaan' memiliki rentang nilai yang sangat bervariasi. Untuk mencegah fitur dengan skala yang jauh lebih besar mendominasi perhitungan jarak oleh algoritma K-Means, normalisasi data dilakukan. Normalisasi ini mentransformasi nilai kekayaan ke dalam rentang yang seragam, biasanya antara 0 dan 1, memastikan bahwa semua fitur memiliki bobot yang setara dalam perhitungan jarak. Implementasi ini dilakukan menggunakan modul *MinMaxScaler* dari *library* *scikit-learn* di Python.
- Pengodean Kategori Bidang Usaha (Label Encoding): Variabel 'Bidang Usaha Utama' merupakan data kategorial (misalnya, "Properti", "Pertambangan", "Manufaktur"). Algoritma K-Means bekerja optimal dengan data numerik. Oleh karena itu, data kategorial ini perlu diubah menjadi representasi numerik. Label Encoding adalah metode sederhana yang memberikan nilai numerik unik untuk setiap kategori unik. Misalnya, "Properti" mungkin diubah menjadi 0, "Pertambangan" menjadi 1, dan seterusnya. Proses ini dilakukan menggunakan modul *LabelEncoder* dari *library* *scikit-learn* di Python.

## 2.3. Konsep algoritma K-means

Algoritma K-Means adalah salah satu metode *unsupervised learning* yang digunakan untuk pengelompokan data [5]. Inti dari algoritma ini adalah mempartisi sejumlah observasi ke dalam  $k$  kluster, di mana setiap observasi termasuk ke dalam kluster yang pusat (*centroid*)nya paling dekat. Algoritma ini bekerja secara iteratif untuk meminimalkan jarak antar data dalam satu kluster terhadap *centroid*-nya [6]. Keunggulan K-Means terletak pada kesederhanaan dan efisiensi komputasinya, menjadikannya pilihan yang kuat untuk dataset berukuran besar.

## 2.4. Penentuan Jumlah Kluster Optimal (K)

Menentukan jumlah kluster ( $K$ ) yang optimal adalah langkah kritis dalam implementasi algoritma K-Means, karena K-Means memerlukan parameter  $K$  sebagai input di awal proses [7]. Dua metode heuristik yang paling sering digunakan untuk tujuan ini adalah Elbow Method dan Silhouette Score. Perhitungan dan visualisasi untuk kedua metode ini dilakukan menggunakan Python.

1. Elbow Method: Metode ini melibatkan perhitungan total inersia (jumlah kuadrat jarak setiap titik ke pusat kluster terdekatnya) untuk berbagai nilai  $K$ . Implementasi K-Means dari *library* *scikit-learn* (*sklearn.cluster.KMeans*) digunakan untuk menghitung inersia. Sebuah grafik yang menampilkan inersia terhadap  $K$  kemudian digambarkan menggunakan *library* *Matplotlib* atau *Seaborn*. Titik di mana kurva grafik menunjukkan "siku" yang jelas, yaitu penurunan inersia mulai melambat secara signifikan, dianggap sebagai nilai  $K$  yang optimal [8].
2. Silhouette Score: Metode ini memberikan metrik kuantitatif untuk mengevaluasi seberapa baik sebuah objek cocok dengan kluster sendiri dan seberapa buruk cocok dengan kluster tetangga [9]. Skor siluet berkisar antara -1 hingga 1; nilai mendekati 1 menunjukkan klusterisasi yang padat dan terpisah dengan baik. Perhitungan Silhouette Score untuk setiap nilai  $K$  dilakukan menggunakan fungsi *silhouette\_score* dari modul *sklearn.metrics* di *Python*. Nilai  $K$  yang menghasilkan *Silhouette Score* rata-rata tertinggi dipilih sebagai optimal

## 2.5. Implementasi Klusterisasi K-Means

Setelah jumlah kluster optimal ( $K$ ) ditentukan, algoritma K-Means diimplementasikan menggunakan *library* *scikit-learn* di *Python* melalui langkah-langkah iteratif berikut [6]:

1. Inisialisasi Centroid: Objek *KMeans* diinisialisasi dengan parameter *n\_clusters* (sesuai  $K$  optimal) dan *random\_state* untuk reproduktivitas. Metode inisialisasi default (*k-means++*) yang lebih cerdas dari

inisialisasi acak murni digunakan untuk mempercepat konvergensi dan mengurangi risiko terjebak pada *local optima*.

2. Penugasan Data ke Kluster: Metode `.fit()` dari objek KMeans dipanggil pada data yang telah diproses. Secara internal, algoritma akan secara iteratif menugaskan setiap titik data ke kluster yang *centroid*-nya memiliki jarak Euclidean terdekat.
3. Pembaruan Posisi Centroid: Setelah penugasan awal, *centroid* dari setiap kluster diperbarui dengan menghitung rata-rata dari semua titik data yang ditugaskan ke kluster tersebut. Langkah ini diulang hingga konvergensi.

Proses Iterasi dan Konvergensi: Algoritma secara otomatis mengulang langkah penugasan dan pembaruan *centroid* hingga posisi *centroid* tidak lagi berubah secara signifikan atau jumlah iterasi maksimum tercapai. Kondisi ini menandakan konvergensi algoritma. Hasil akhir berupa label kluster untuk setiap titik data dan koordinat *centroid* kluster dapat diakses melalui atribut objek KMeans

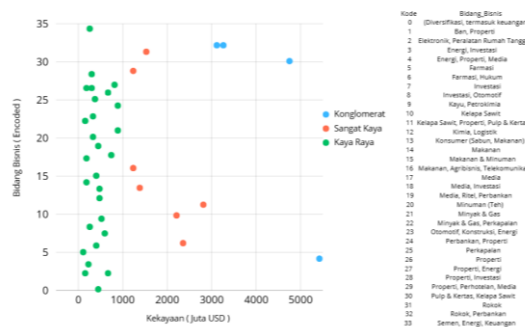
### 3. Hasil dan Diskusi



#### 3.1. Penentuan Jumlah Kluster Optimal ( $K$ )

Hasil Dalam proses penentuan jumlah kluster optimal, digunakan dua metode evaluasi: *Silhouette Score* dan *Elbow Method*. Gambar berikut menunjukkan hasil dari masing-masing metode untuk jumlah kluster  $K$  dari 2 hingga 9. Grafik sebelah kiri menunjukkan nilai *Silhouette Score*, yang mengukur seberapa baik suatu titik data cocok dengan klusternya dibandingkan dengan kluster lain. Nilai tertinggi terdapat pada  $K=2$ , yaitu sekitar 0.63, diikuti oleh nilai stabil di sekitar 0.53–0.54 pada  $K=4$  hingga  $K=6$ . Penurunan nilai terjadi secara konsisten setelah  $K=6$ . Meskipun  $K=2$  memberikan skor tertinggi, nilai tersebut kurang mencerminkan segmentasi yang realistis dalam konteks kekayaan ekonomi yang lebih kompleks. Oleh karena itu, perlu dipertimbangkan keseimbangan dengan metode lain. Grafik sebelah kanan merupakan visualisasi *Elbow Method*, yang menunjukkan nilai inerti—yaitu jumlah kuadrat jarak antar titik ke pusat kluster. Terlihat penurunan tajam dari  $K=2$  ke  $K=3$ , lalu penurunan mulai melambat dari  $K=4$  ke atas. Titik siku (elbow) yang terlihat jelas pada  $K=3$  menunjukkan bahwa penambahan kluster setelah titik tersebut tidak lagi memberikan pengurangan inerti yang signifikan. Berdasarkan kedua grafik tersebut, jumlah kluster optimal yang paling representatif adalah  $K=3$ . Nilai ini merupakan titik kompromi antara kohesi internal kluster dan kompleksitas model. Oleh karena itu, analisis selanjutnya menggunakan  $K=3$  sebagai jumlah kluster dalam proses K-Means.

#### 3.2. Deskripsi kluster



Gambar 1. Grafik Klusterisasi Kekayaan Berdasarkan Bidang Bisnis

DOI: <https://doi.org/10.31004/riggs.v4i2.1382>

Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

Dari grafik dapat diamati bahwa Klaster “Kaya Raya” (hijau) mendominasi individu dengan kekayaan rendah dan tersebar luas pada berbagai bidang bisnis. Ini menunjukkan bahwa kelompok ini relatif lebih banyak jumlahnya dan aktif di sektor yang lebih beragam, mulai dari manufaktur, makanan, hingga properti.

Klaster “Sangat Kaya” (oranye) cenderung menempati posisi menengah pada sumbu kekayaan dan terlihat lebih terfokus pada bidang bisnis tertentu, seperti energi, kelapa sawit, atau media. Meskipun kekayaan mereka tidak setinggi kelompok konglomerat, konsistensi dalam satu sektor menunjukkan spesialisasi usaha.

Klaster “Konglomerat” (biru) sangat menonjol di bagian kanan grafik karena kekayaannya yang jauh lebih besar dari kelompok lain. Jumlah individu dalam klaster ini sedikit, namun posisi kekayaan mereka jauh di atas rata-rata. Mereka umumnya aktif di sektor strategis dan terdiversifikasi, seperti perbankan, rokok, dan properti, yang menjadi sumber stabil penghasilan besar. Pola ini menunjukkan bahwa diversifikasi lintas sektor menjadi ciri khas dari kelompok konglomerat.

Secara umum, visualisasi ini menunjukkan bahwa terdapat keterkaitan antara besarnya kekayaan dan jenis bidang usaha, di mana semakin tinggi kekayaan, cenderung semakin strategis dan terarah bidang bisnisnya. Pola sebaran ini memperkuat keabsahan hasil klasterisasi dan segmentasi kelas kekayaan.

### 3.3 Penentuan Jumlah Klaster Optimal ( $K$ )

Hasil Dalam proses penentuan jumlah klaster optimal, digunakan dua metode evaluasi: *Silhouette Score* dan *Elbow Method*. Gambar berikut menunjukkan hasil dari masing-masing metode untuk jumlah klaster  $K$  dari 2 hingga 9.

Grafik sebelah kiri menunjukkan nilai *Silhouette Score*, yang mengukur seberapa baik suatu titik data cocok dengan klasternya dibandingkan dengan klaster lain. Nilai tertinggi terdapat pada  $K=2$ , yaitu sekitar 0.63, diikuti oleh nilai stabil di sekitar 0.53–0.54 pada  $K=4$  hingga  $K=6$ . Penurunan nilai terjadi secara konsisten setelah  $K=6$ . Meskipun  $K=2$  memberikan skor tertinggi, nilai tersebut kurang mencerminkan segmentasi yang realistis dalam konteks kekayaan ekonomi yang lebih kompleks. Oleh karena itu, perlu dipertimbangkan keseimbangan dengan metode lain.

### 3.2. Deskripsi klaster

Hasil klasterisasi yang divisualisasikan pada Gambar 3 menunjukkan hubungan yang menarik antara tingkat kekayaan dan bidang usaha para individu dalam daftar orang terkaya tahun 2007. Pola distribusi yang terbentuk mencerminkan perbedaan karakteristik yang cukup tajam antara tiga kategori kekayaan yang telah didefinisikan secara manual.

Klaster “Konglomerat” menunjukkan konsentrasi individu yang memiliki kekayaan sangat tinggi dan umumnya beroperasi di sektor-sektor ekonomi strategis seperti perbankan, energi, dan infrastruktur. Keberadaan mereka di sisi kanan grafik, dengan nilai kekayaan jauh di atas klaster lain, menandakan bahwa model bisnis mereka kemungkinan besar bersifat diversifikasi lintas sektor. Diversifikasi ini dapat berfungsi sebagai strategi mitigasi risiko dan pengoptimalan pendapatan jangka panjang. Hal ini konsisten dengan teori bahwa semakin kompleks portofolio usaha, semakin tinggi pula akumulasi kekayaan yang mungkin dicapai.

Klaster “Sangat Kaya” terdiri dari individu dengan kekayaan menengah yang cenderung fokus pada satu atau dua sektor dominan, seperti kelapa sawit, properti, atau manufaktur berat. Spesialisasi ini mencerminkan kedalaman dan penguasaan terhadap bidang tertentu, meskipun dengan risiko yang lebih tinggi terhadap gejolak sektor tunggal.

Sementara itu, klaster “Kaya Raya” dihuni oleh individu yang kekayaannya masih berada di bawah 1000 juta USD dan tersebar di berbagai bidang bisnis, termasuk konsumsi, makanan, dan produk rumah tangga. Pola penyebaran titik yang lebih luas secara vertikal menunjukkan bahwa kelompok ini lebih banyak secara jumlah dan menjalankan usaha di sektor yang relatif lebih umum. Mereka memiliki potensi pertumbuhan yang besar namun belum mencapai skala ekonomi yang sama dengan dua klaster lainnya.

Dengan demikian, hasil ini menunjukkan bahwa tingkat diversifikasi dan skala sektor usaha memiliki keterkaitan yang erat dengan besaran kekayaan, dan klasterisasi berbasis data mampu mengungkap segmentasi tersebut secara objektif

#### 4. Kesimpulan

Penelitian ini berhasil mengelompokkan individu dalam daftar orang terkaya di Indonesia tahun 2007 ke dalam tiga kluster menggunakan algoritma K-Means, yang memberikan wawasan penting mengenai distribusi kekayaan dan keterkaitannya dengan karakteristik sektor usaha dominan. Algoritma K-Means terbukti efektif untuk eksplorasi data ekonomi, mampu mengidentifikasi pola tersembunyi dalam dataset kekayaan, dan dapat menjadi alat bantu yang berharga dalam analisis ekonomi serta perumusan strategi investasi atau kebijakan. Untuk pengembangan penelitian di masa mendatang, disarankan untuk melakukan analisis klusterisasi dengan data dari tahun-tahun berbeda guna melihat tren distribusi kekayaan dan pergeseran pola klusterisasi seiring waktu, serta dampaknya terhadap ekonomi nasional. Selain itu, integrasi variabel tambahan seperti usia pengusaha, lokasi geografis operasional bisnis, atau tingkat investasi asing dapat memberikan analisis yang lebih komprehensif dan multidimensional, atau bahkan membandingkan metode klusterisasi lain seperti DBSCAN atau hierarchical clustering sebagai pembanding untuk mengevaluasi konsistensi dan robustnya hasil

#### Referensi

- [1] M. H. Siregar, R. B. A., & R. Efendi, "Studi Konsentrasi Kekayaan dan Implikasinya Terhadap Perekonomian Nasional," *Jurnal Ekonomi Pembangunan*, vol. 15, no. 1, hal. 45-60, 2018.
- [2] L. N. Anggraini, D. H. Putra, dan B. Santoso, "Analisis Kesenjangan Kekayaan di Indonesia Menggunakan Gini Ratio," *Jurnal Studi Ekonomi Indonesia*, vol. 12, no. 2, hal. 112-125, 2019.
- [3] S. Wibowo, *Metode Penelitian untuk Bisnis dan Ekonomi*. Jakarta: Salemba Empat, 2019.
- [4] N. R. Khasanah dan M. Hidayat, "Pra-Pemrosesan Data untuk Peningkatan Akurasi Klasifikasi pada Data Tidak Seimbang," *Jurnal Teknologi Informasi Komunikasi*, vol. 6, no. 1, hal. 25-34, 2019.
- [5] E. T. Wibowo, "Implementasi Algoritma K-Means Clustering untuk Pengelompokan Data Mahasiswa," *Jurnal Informatika*, vol. 12, no. 1, hal. 1-8, 2018.
- [6] D. Setiadi dan E. Yulianti, "Penerapan Algoritma K-Means dalam Klusterisasi Data Penjualan Produk," *Jurnal Rekayasa Sistem Industri*, vol. 7, no. 2, hal. 110-120, 2019.
- [7] B. Prasetyo, *Penerapan Data Mining: Konsep dan Aplikasi*. Yogyakarta: Andi, 2012.
- [8] A. Fitri dan R. K. Dewi, "Optimalisasi Jumlah Kluster Menggunakan Metode Elbow dan Silhouette pada Data Kependudukan," *Jurnal Rekayasa Perangkat Lunak*, vol. 7, no. 1, hal. 1-10, 2020.
- [9] S. Wijaya dan R. Lestari, "Validasi Hasil Klusterisasi Menggunakan Indeks Davies-Bouldin dan Silhouette Coefficient," *Jurnal Sains Komputer dan Informatika*, vol. 4, no. 2, hal. 70-78, 2020.
- [10] B. Jatmiko dan K. B. Adi, "Analisis Perbandingan Algoritma Klustering K-Means dan Hierarchical Clustering pada Data Penjualan," *Jurnal Sistem Informasi Bisnis*, vol. 10, no. 1, hal. 1-8, 2020.
- [11] M. D. Oktaviani dan R. A. Sukma, "Pengelompokan Jenis Pelayanan Kesehatan Menggunakan K-Means Clustering," *Jurnal Ilmiah Komputer dan Informatika*, vol. 7, no. 2, hal. 130-138, 2021.
- [12] H. P. Sari, A. Nugroho, dan S. H. Syahputra, "Penerapan Algoritma K-Means untuk Klusterisasi Data Potensi Wisata Alam," *Jurnal Informatika Ekonomi Bisnis*, vol. 5, no. 1, hal. 10-18, 2020.