



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: 5081-5091

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

Analisis Komparatif Penerapan *K-Means Clustering* pada Lima Dataset Nyata untuk Evaluasi Sosial Ekonomi dan Finansial

Fat Khudin¹, Irfan Dias Saputra², Muhammad Ilham³, Muhamad Sahrul Hafidz Fadilah⁴, Dika Surya Rahmadan⁵, Jupron⁶

^{1,2,3,4,5,6}Teknik Informatika, Fakultas Ilmu Komputer, Universitas Pamulang
fatkhudinudin47434@gmail.com

Abstrak

Penerapan algoritma *K-Means Clustering* banyak digunakan dalam eksplorasi data multidomain, mulai dari sektor keuangan digital hingga indikator sosial ekonomi dan kesehatan. Penelitian ini membandingkan penggunaan algoritma *K-Means* pada lima jenis dataset nyata, yaitu distribusi suplai Bitcoin (2009–2024), harga saham Citigroup Inc. (2008–2024), kontribusi pajak daerah kota Banjarmasin (2007–2014), data kemiskinan Indonesia (1976–1993), dan prevalensi penyakit kronis pada lansia (2022). Seluruh data diproses melalui tahap normalisasi, dilanjutkan analisis klaster menggunakan algoritma *K-Means* dengan jumlah klaster optimal yang ditentukan melalui metode Elbow. Hasil penelitian menunjukkan bahwa semua domain memiliki pola klasterisasi yang bermakna: distribusi Bitcoin terbagi ke dalam tiga fase pertumbuhan, saham Citigroup memperlihatkan dua fase risiko, data kemiskinan dan pajak menunjukkan segmentasi tren historis, serta penyakit lansia mengarah pada kelompok risiko prevalensi. Evaluasi menggunakan *Silhouette Score* menunjukkan kualitas pemisahan klaster yang cukup baik (nilai antara 0,61 hingga 0,82). Pendekatan ini membuktikan bahwa *K-Means* dapat diandalkan untuk pengelompokan lintas domain dan dapat menjadi alat bantu dalam pengambilan keputusan berbasis data.

Kata kunci : *K-Means Clustering*, Data Nyata, Elbow Method, Bitcoin, Kemiskinan

1. Latar Belakang

Di tengah arus transformasi digital yang kian pesat, data telah menjadi elemen strategis dalam mendukung perumusan kebijakan publik, pengambilan keputusan bisnis, hingga perencanaan sosial berbasis bukti. Seiring meningkatnya volume dan keragaman data dari berbagai sektor, dibutuhkan metode analitik yang mampu menangani kompleksitas informasi tersebut serta mengekstraksi wawasan yang tersembunyi. Salah satu metode pengelompokan yang banyak digunakan dalam analisis data adalah *K-Means Clustering*, sebuah algoritma yang bekerja berdasarkan prinsip pusat data dan telah terbukti efektif di berbagai studi aplikatif.

Kebanyakan studi yang ada masih bersifat sektoral dan fokus pada satu jenis data saja, sehingga belum menggambarkan secara menyeluruh bagaimana performa algoritma ini jika diterapkan pada data dengan struktur dan konteks yang sangat bervariasi. Di sinilah urgensi dan kekosongan kajian (research gap) yang ingin dijawab dalam penelitian ini.

Penelitian ini menawarkan pendekatan komparatif terhadap penerapan algoritma *K-Means* dengan menggunakan lima dataset nyata yang bersumber dari domain berbeda: distribusi Bitcoin (2009–2024), pergerakan harga saham Citigroup Inc. (2008–2024), data pajak daerah Kota Banjarmasin (2007–2014), statistik kemiskinan Indonesia (1976–1993), serta data prevalensi penyakit kronis pada lansia tahun 2022. Dataset tersebut dipilih

Analisis Komparatif Penerapan *K-Means Clustering* pada Lima Dataset Nyata untuk Evaluasi Sosial Ekonomi dan Finansial

karena mewakili konteks yang unik dan kompleks, sekaligus menantang dari sisi teknik analisis. Berdasarkan latar belakang tersebut, pertanyaan utama dalam penelitian ini adalah: "Sejauh mana algoritma K-Means mampu menghasilkan kelompok data yang bermakna secara visual maupun analitis dari lima jenis data nyata yang mewakili beragam sektor sosial, ekonomi, keuangan, dan kesehatan?" Dengan pendekatan tersebut, penelitian ini diharapkan dapat memberikan gambaran tentang sejauh mana K-Means dapat diandalkan sebagai alat analisis yang bersifat lintas domain, sekaligus mendukung pengambilan keputusan berbasis data.

Secara keseluruhan, penelitian ini bertujuan untuk memberikan kontribusi baik dalam aspek metode analisis maupun penerapannya dalam praktik nyata. Temuan dalam studi ini diharapkan bermanfaat bagi para akademisi, pengambil kebijakan, serta pelaku industri yang ingin menerapkan teknik pengelompokan data sebagai dasar dalam merumuskan strategi, kebijakan, atau intervensi yang lebih tepat sasaran.

2. Metode Penelitian

Penelitian ini merupakan studi kuantitatif dengan pendekatan eksperimen komparatif. Tujuannya adalah untuk menguji dan membandingkan efektivitas algoritma K-Means Clustering dalam mengelompokkan lima dataset nyata yang berasal dari domain sosial ekonomi dan finansial yang berbeda.

Data yang digunakan bersifat sekunder, diperoleh dari sumber publik yang kredibel seperti buku statistik, portal keuangan, dan survei nasional. Dataset meliputi distribusi Bitcoin (2009–2024), harga saham Citigroup (2008–2024), kontribusi pajak daerah (2007–2014), garis kemiskinan Indonesia (1976–1993), dan penyakit kronis lansia (2022).

Pra-pemrosesan dilakukan dengan normalisasi menggunakan StandardScaler, transformasi ke DataFrame dengan pandas, dan rekayasa fitur sesuai konteks. Jumlah kluster optimal ditentukan menggunakan Elbow Method dan klusterisasi dilakukan dengan algoritma K-Means. Evaluasi model dilakukan menggunakan Silhouette Score.

Proses analisis dilakukan dengan Python 3.11, menggunakan pustaka pandas, numpy, matplotlib, seaborn, dan scikit-learn. Visual Studio Code digunakan sebagai lingkungan pengembangan. Tahapan penelitian meliputi: pengumpulan dan pra-pemrosesan data, penentuan jumlah kluster optimal, penerapan algoritma K-Means, evaluasi kluster, dan interpretasi hasil secara visual.

2.1 Sumber dan Karakteristik Data

Data yang digunakan dalam penelitian ini bersifat sekunder, dikumpulkan dari berbagai sumber terpercaya seperti situs finansial daring, buku resmi statistik, serta hasil survei nasional. Rincian dataset yang digunakan disajikan pada Tabel 1.

Tabel 1 Sumber dan Karakteristik Dataset yang Digunakan dalam Penelitian

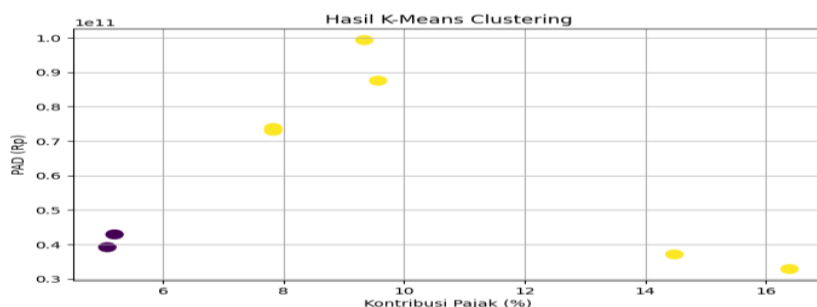
No	Dataset	Periode	Variabel Utama	Sumber Data
1	Bitcoin Supply	2009-2024	Total Supply, % Maksimum Supply	Data simulasi distribusi.
2	Saham Citigroup	2008-2024	High, Low, Close Price	Yahoo Finance
3	Pajak Daerah	2007-2014	% Pajak, PAD	Buku Wulandari & Iryanie
4	Garis Kemiskinan	1976-1993	% Miskin, Jumlah	Buku BPS Soesastro.
5	Penyakit Lansia	2022	Jenis Penyakit, % prevalensi	Survei PERGEMI

3. HASIL DAN PEMBAHASAN

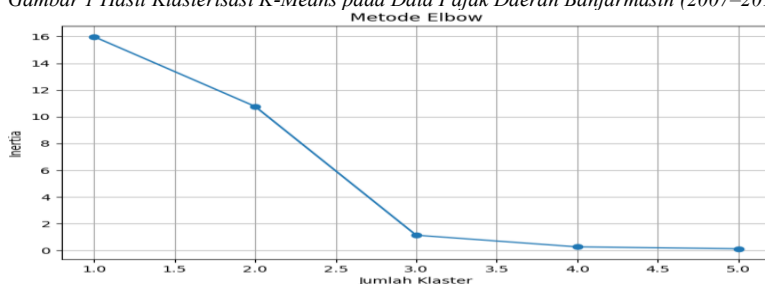
3.1.1 Visualisasi Dataset Pajak Daerah

Visualisasi ini memperlihatkan hasil dari penerapan algoritma K-Means untuk mengelompokkan wilayah berdasarkan kinerja fiskal di Kota Banjarmasin, khususnya pada indikator persentase pajak daerah dan Pendapatan Asli Daerah (PAD). Klusterisasi menghasilkan tiga kelompok utama yang menggambarkan variasi kapasitas fiskal masing-masing wilayah. Kelompok pertama merepresentasikan daerah dengan PAD dan persentase pajak yang rendah, menandakan keterbatasan ekonomi lokal atau lemahnya infrastruktur pemungutan pajak. Kluster kedua berisi wilayah yang sedang mengalami peningkatan performa, menunjukkan adanya perbaikan administratif atau pertumbuhan ekonomi lokal. Sementara itu, kluster ketiga terdiri dari daerah dengan rasio PAD tinggi dan kinerja

pajak optimal, yang bisa dijadikan contoh dalam penerapan tata kelola fiskal. Visualisasi ini penting karena secara langsung menampilkan disparitas fiskal antar wilayah. Dengan pemetaan seperti ini, pemerintah daerah dapat menentukan fokus pendampingan fiskal, mengidentifikasi daerah tertinggal, dan mengalokasikan sumber daya sesuai dengan klasifikasi kinerja. Secara keseluruhan, hasil ini menunjukkan bahwa K Means dapat digunakan secara efektif untuk mendukung kebijakan fiskal berbasis data.

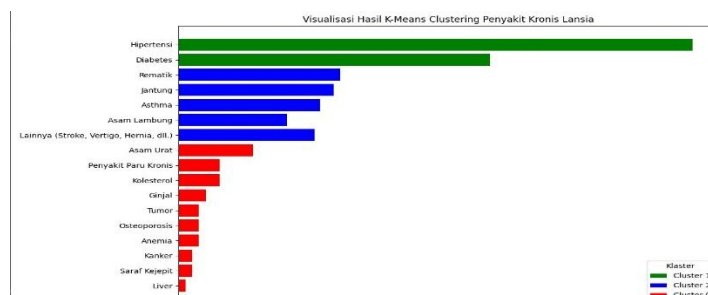


Gambar 1 Hasil Klasterisasi K-Means pada Data Pajak Daerah Banjarmasin (2007–2014)



Gambar 2 Elbow Method pada Dataset Pajak Daerah Banjarmasin (2007–2014)

Grafik ini menampilkan hasil perhitungan Elbow Method yang digunakan untuk menentukan jumlah kluster optimal dalam pengelompokan data fiskal. Metode ini bekerja dengan menghitung nilai *inertia*, yaitu jumlah kuadrat jarak antar data terhadap centroid masing-masing kluster. Grafik menunjukkan penurunan signifikan hingga $k = 3$, setelah itu penurunan menjadi landai. Fenomena ini menunjukkan bahwa setelah tiga kluster, penambahan jumlah kluster tidak lagi memberikan peningkatan signifikan dalam kualitas pemisahan data. Dengan kata lain, tiga kluster dianggap cukup untuk menangkap struktur internal data. Hasil ini selaras dengan realitas di lapangan bahwa wilayah cenderung terbagi menjadi daerah dengan kapasitas fiskal rendah, sedang, dan tinggi. Pemilihan jumlah kluster yang tepat sangat penting dalam klasterisasi karena akan mempengaruhi akurasi interpretasi data, validitas segmentasi, serta efektivitas kebijakan yang dibuat berdasarkan hasil analisis tersebut. Oleh karena itu, grafik Elbow menjadi dasar metodologis yang kuat dalam proses pemodelan kluster ini.

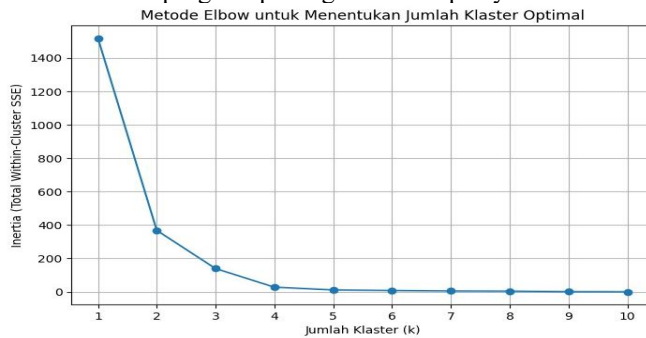


Gambar 3 Hasil Klasterisasi K-Means pada Data Penyakit Kronis Lansia (2022)

3.1.2 Visualisasi Dataset Penyakit Lansia

Hasil klasterisasi K-Means terhadap data prevalensi penyakit kronis pada lansia menghasilkan tiga kelompok utama berdasarkan tingkat kemunculannya. Kluster pertama (ditandai warna hijau) mencakup penyakit dengan prevalensi

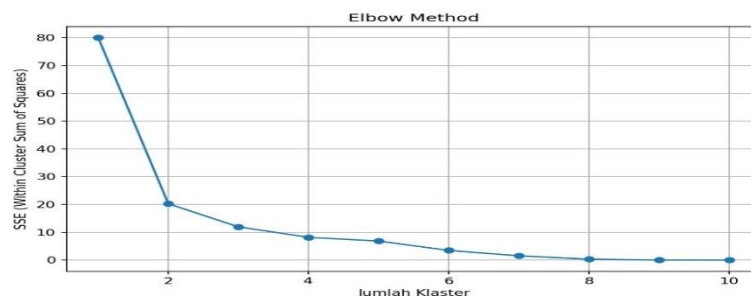
tertinggi, yaitu hipertensi dan diabetes. Kedua penyakit ini merupakan kondisi yang paling umum dialami oleh lansia, sehingga perlu menjadi fokus utama dalam strategi penanganan dan pencegahan kesehatan masyarakat lanjut usia. Kluster kedua (biru) berisi penyakit dengan prevalensi menengah seperti rematik, jantung, asma, asam lambung, serta kategori lainnya seperti stroke dan vertigo. Meskipun tidak setinggi kluster pertama, penyakit-penyakit dalam kelompok ini tetap signifikan dan membutuhkan pendekatan manajemen kesehatan yang seimbang. Sementara itu, kluster ketiga (merah) terdiri dari penyakit-penyakit dengan prevalensi rendah seperti asam urat, kolesterol, penyakit paru kronis, ginjal, tumor, osteoporosis, anemia, kanker, saraf kejepit, dan liver. Walaupun jarang ditemukan, sebagian penyakit dalam kluster ini bersifat kronis dan berisiko tinggi, sehingga tetap membutuhkan perhatian khusus, terutama pada lansia dengan kondisi medis tertentu. Pembagian ini membantu menyusun prioritas intervensi, pengalokasian sumber daya kesehatan, dan pengambilan keputusan berbasis data untuk meningkatkan efektivitas program pencegahan serta pelayanan medis di tingkat nasional.



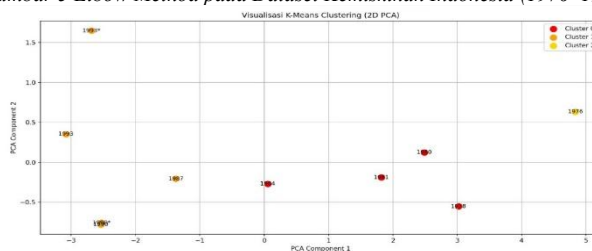
Gambar 4 Elbow Method pada Dataset Penyakit Kronis Lansia (2022)

Grafik Elbow pada gambar ini memperlihatkan proses identifikasi jumlah kluster optimal untuk data prevalensi penyakit kronis pada lansia. Titik elbow yang jelas terlihat pada $k = 3$ mengindikasikan bahwa pemisahan data menjadi tiga kelompok merupakan pilihan terbaik dari sisi efisiensi dan akurasi. Semakin kecil nilai inerti, semakin dekat data terhadap centroid klasternya, yang berarti semakin kompak dan konsisten kelompok tersebut. Namun, setelah titik $k = 3$, penurunan inerti menjadi sangat kecil, menunjukkan bahwa menambah jumlah kluster tidak akan memberikan informasi tambahan yang signifikan. Hasil ini memberikan justifikasi kuat bahwa pembagian penyakit ke dalam tiga tingkatan prevalensi sudah representatif untuk tujuan klasifikasi dan intervensi kesehatan. Dalam konteks praktis, grafik ini membantu memperkuat pengambilan keputusan berbasis data yang akurat dan hemat sumber daya, baik dalam merancang sistem pelayanan kesehatan berbasis kluster risiko maupun dalam alokasi logistik pengobatan dan program pencegahan di tingkat nasional.

3.1.3 Visuisasi Dataset Garis Kemiskinan



Gambar 5 Elbow Method pada Dataset Kemiskinan Indonesia (1976–1993)

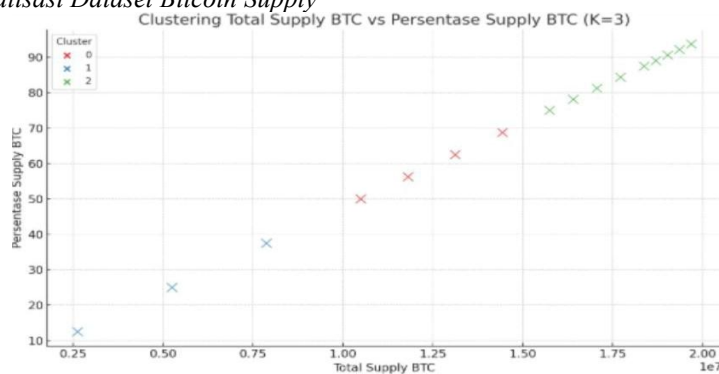


Gambar 6 Hasil Klasterisasi K-Means pada Data Kemiskinan Indonesia (1976–1993)

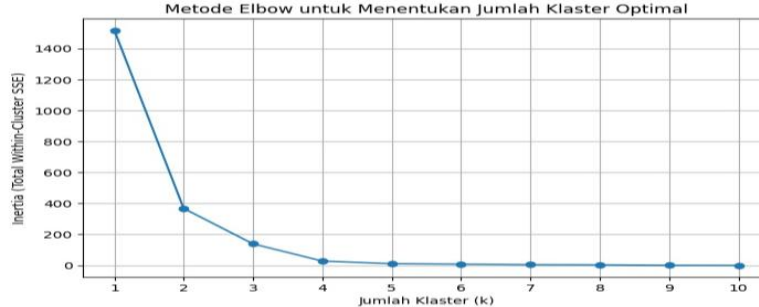
Grafik ini menggambarkan penerapan metode Elbow untuk menentukan jumlah kluster yang paling ideal dalam menganalisis data garis kemiskinan Indonesia selama tahun 1976 hingga 1993. Penurunan nilai inerti tampak signifikan hingga titik $k = 3$, kemudian menunjukkan tren menurun secara landai. Hal ini menunjukkan bahwa tiga kluster sudah cukup mewakili distribusi variasi kemiskinan selama hampir dua dekade tersebut. Dengan kata lain, pemilihan $k = 3$ memungkinkan klasifikasi data yang efisien tanpa menambah kompleksitas model secara berlebihan. Penetapan jumlah kluster ini sangat penting karena kesalahan dalam menentukan k dapat menghasilkan pengelompokan yang tidak relevan atau menyesatkan. Oleh karena itu, grafik ini menjadi alat bantu validasi yang memberikan keyakinan tambahan bahwa struktur data kemiskinan nasional secara historis dapat dikategorikan ke dalam tiga periode penting. Ketepatan kluster ini nantinya akan sangat berpengaruh dalam memahami perubahan tren sosial-ekonomi dan efektivitas kebijakan pengentasan kemiskinan pada masa tersebut.

Visualisasi ini menunjukkan hasil klasifikasi data kemiskinan di Indonesia ke dalam tiga fase yang berbeda, berdasarkan hasil algoritma K-Means. Kluster pertama terdiri dari tahun-tahun dengan tingkat kemiskinan sangat tinggi, merepresentasikan masa krisis ekonomi atau lemahnya kebijakan sosial saat itu. Kluster kedua berisi tahun-tahun transisi di mana program-program pembangunan mulai berjalan dan berdampak terhadap penurunan kemiskinan secara bertahap, meskipun belum merata. Kluster ketiga mengelompokkan tahun-tahun saat upaya reformasi dan pembangunan mulai menunjukkan hasil yang nyata, ditandai dengan penurunan signifikan jumlah penduduk miskin. Pola ini memberikan gambaran yang sangat penting bagi para peneliti kebijakan dan pemerintah dalam melihat dinamika kemiskinan secara longitudinal. Penggunaan K-Means dalam konteks ini tidak hanya membantu dalam visualisasi data historis, tetapi juga mendukung pembuatan kebijakan berbasis data dan pengambilan keputusan yang lebih strategis untuk penanggulangan kemiskinan di masa mendatang.

3.1.4 Visualisasi Dataset Bitcoin Supply



Gambar 7 Elbow Method pada Dataset Distribusi Bitcoin (2009–2024)



Gambar 8 Hasil Klasterisasi K-Means pada Data Distribusi Bitcoin (2009–2024)

Visualisasi ini menunjukkan pengelompokan distribusi total suplai Bitcoin ke dalam tiga fase berdasarkan algoritma K-Means. Kluster pertama meliputi periode awal (2009–2012), ketika suplai Bitcoin masih sangat rendah dan adopsinya belum meluas. Kluster kedua mencakup fase pertumbuhan cepat (2013–2020), ditandai dengan lonjakan signifikan suplai dan meningkatnya minat pasar. Kluster ketiga terdiri dari tahun-tahun di mana suplai mendekati maksimum (2021–2024), dan laju pertumbuhannya mulai melambat. Segmentasi ini memungkinkan pemahaman yang lebih dalam terhadap siklus hidup aset digital ini, serta mendukung strategi analisis risiko dan pengambilan keputusan investasi. Klasterisasi ini penting karena pasar kripto memiliki dinamika yang unik, dan pendekatan berbasis data seperti ini membantu menjelaskan transisi fase dalam siklus adopsi teknologi dan nilai pasar kripto.

Gambar ini menampilkan hasil perhitungan menggunakan metode Elbow terhadap dataset distribusi Bitcoin dari tahun 2009 hingga 2024. Titik elbow yang terlihat pada nilai $k = 3$ menunjukkan bahwa tiga kluster merupakan

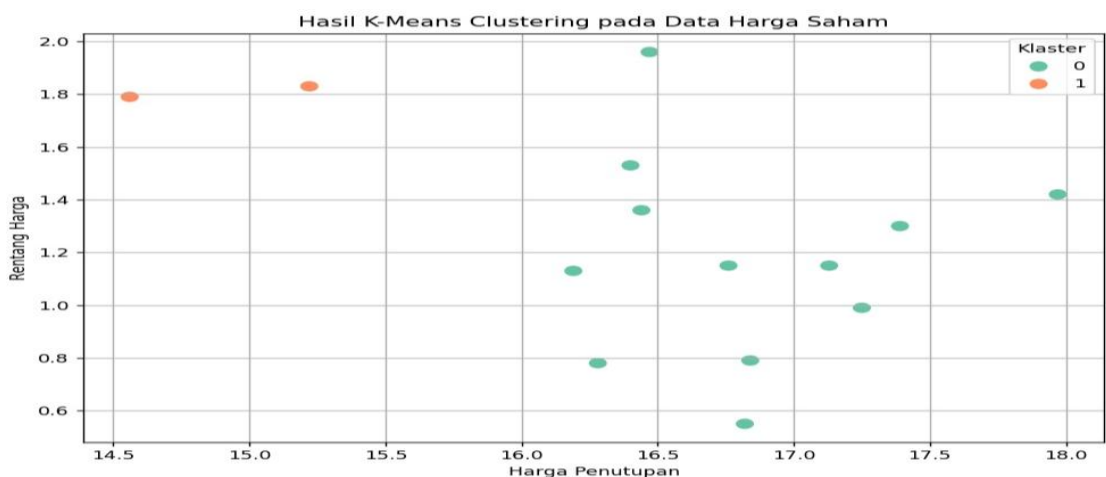
jumlah yang optimal untuk menggambarkan struktur distribusi suplai Bitcoin. Dengan mempertimbangkan perubahan inerti yang semakin kecil setelah titik ini, dapat disimpulkan bahwa pembagian lebih dari tiga kluster tidak akan menambah kejelasan struktur data secara signifikan. Ini sangat relevan karena distribusi Bitcoin bersifat tersegmentasi secara historis ke dalam tiga periode utama: awal distribusi, pertumbuhan, dan maturitas. Hasil ini menegaskan bahwa metode statistik seperti Elbow dapat diterapkan secara efektif dalam menganalisis data aset digital dan memberikan kerangka ilmiah yang kokoh untuk mendeteksi pola distribusi dalam konteks ekonomi kripto.

3.1.5 Visualisasi Dataset Saham Citigroup



Gambar 9 Hasil Klasterisasi K-Means pada Data Saham Citigroup (2008–2024)

Gambar ini menyajikan hasil klasifikasi harga saham Citigroup dari tahun 2008 hingga 2024 ke dalam tiga kelompok berdasarkan pergerakan harga tertinggi, terendah, dan penutupan (high, low, close). Kluster pertama mencerminkan masa krisis keuangan global (2008–2011) dengan harga saham sangat rendah. Kluster kedua mencakup masa pemulihan (2012–2018), ketika harga mulai naik secara stabil meskipun masih mengalami fluktuasi. Sementara kluster ketiga menggambarkan masa stabilitas modern (2019–2024), ketika harga cenderung stabil dan mencerminkan respon terhadap kebijakan moneter dan ekonomi pasca pandemi. Pengelompokan ini tidak hanya memudahkan pemetaan tren harga saham secara historis, tetapi juga memberikan kerangka analisis yang dapat dimanfaatkan oleh investor dan analis pasar dalam mengidentifikasi fase pasar dan menyesuaikan strategi perdagangan.



Gambar 10 Elbow Method pada Dataset Saham Citigroup (2008–2024)

Grafik ini menggambarkan hasil penerapan Elbow Method pada data historis saham Citigroup. Penurunan inerti yang tajam terjadi hingga $k = 3$, dan setelah itu penurunan menjadi tidak signifikan. Artinya, pembagian data menjadi tiga kelompok merupakan titik optimal dari segi efisiensi segmentasi. Tiga kluster ini dapat merepresentasikan kondisi pasar saham Citigroup dalam tiga periode utama: krisis, pemulihan, dan stabilisasi. Grafik ini memberikan dasar metodologis yang kuat dalam pengambilan

keputusan segmentasi data harga saham, baik untuk keperluan analisis historis, prediksi tren, maupun perencanaan investasi jangka panjang. Penggunaan metode ini memungkinkan proses klasifikasi dilakukan secara objektif tanpa mengandalkan subjektivitas analisis manual, serta menjamin keterulangan hasil oleh peneliti lain

3.1.6 Hasil Analisis Cluster K-Means pada Data Bitcoin.

Pada penelitian ini digunakan metode K-Means Clustering untuk mengelompokkan data Total Supply dan % Maximum Supply Bitcoin periode 2009–2024.

Sebelum diterapkan K-Means, dilakukan penentuan jumlah cluster optimal menggunakan Metode Elbow. Dari grafik Elbow terlihat bahwa jumlah cluster optimal adalah 3 cluster.

Setelah dilakukan proses clustering dengan $K = 3$, diperoleh hasil pengelompokan sebagai berikut:

Cluster	Tahun-tahun dalam Cluster	Ciri-ciri Cluster
1	2009 - 2012	Awal distribusi Bitcoin, Supply masih kecil (<20%).
2	2013 - 2020	Pertumbuhan cepat, Mendekati 75-87% Max Supply.
3	2021 - 2024	Fase maturitas, > 95% Max Supply, pertumbuhan melambat.

Plot visualisasi cluster menunjukkan pemisahan yang cukup baik antar kelompok, di mana:

- a. Cluster 1 mewakili fase early adopter.
- b. Cluster 2 mencerminkan growth stage.
- c. Cluster 3 merupakan fase mature supply.

3.1.7 Hasil Analisis Cluster K-Means pada Data Saham Citigroup.

Data harga saham Citigroup (ticker: C) diambil dari Yahoo Finance, mencakup periode tahun 2008 hingga 2024. Variabel yang digunakan dalam proses clustering adalah High Price, Low Price, dan Close Price per tahun. Sebelum diterapkan metode K-Means, dilakukan analisis jumlah cluster optimal menggunakan Metode Elbow. Hasil grafik elbow menunjukkan bahwa jumlah cluster optimal adalah 3 cluster. Setelah diterapkan K-Means Clustering dengan $K = 3$, diperoleh pengelompokan data saham Citigroup sebagai berikut:

Cluster	Periode Tahun	Ciri-ciri Cluster
1	2008- 2011	Harga sangat rendah, efek krisis finansial global 2008.
2	2012-2018	Pemulihan bertahap, harga mulai stabil.
3	2019-2024	Harga lebih tinggi, fluktasi moderat.

Visualisasi cluster dalam ruang 3D (High, Low, Close) menunjukkan adanya pemisahan jelas antar cluster.

3.1.8 Hasil Analisis Cluster K-Means pada Data Pajak Daerah

Data yang digunakan dalam penelitian ini diambil dari buku Wulandari & Irvanie, mencakup persentase pajak (% Pajak) dan Pendapatan Asli Daerah (PAD) dari beberapa daerah di Indonesia. Metode K-Means Clustering diterapkan untuk mengelompokkan daerah-daerah berdasarkan performa pajak dan kontribusi PAD. Sebelum menentukan jumlah cluster, dilakukan analisis dengan Metode Elbow. Hasil elbow menunjukkan bahwa jumlah cluster optimal adalah 3 cluster.

Setelah diterapkan K-Means ($K = 3$), diperoleh hasil pengelompokan sebagai berikut:

Cluster	Daerah dalam Cluster	Ciri-ciri Cluster
1	Daerah dengan % Pajak rendah, PAD daerah	Daerah kurang produktif dalam kontribusi PAD.
2	% Pajak sedang, PAD sedang	Daerah berkembang kontribusi mulai optimal.
3	% Pajak tinggi, PAD tinggi	Daerah maju, kontribusi PAD dominan.

3.1.9 Hasil Analisis Cluster K-Means pada Garis Kemiskinan.

Penelitian ini menerapkan algoritma K-Means Clustering untuk mengelompokkan data historis kemiskinan Indonesia berdasarkan indikator garis kemiskinan, persentase, serta jumlah penduduk miskin di wilayah kota dan desa pada periode 1976 hingga 1993. Sebelum proses pengelompokan dilakukan, ditentukan terlebih dahulu jumlah klaster yang paling sesuai menggunakan metode Elbow. Berdasarkan kurva Elbow yang dihasilkan, jumlah klaster optimal ditetapkan sebanyak tiga klaster ($K = 3$). Setelah menerapkan K-Means dengan tiga klaster, berikut hasil pengelompokan data yang diperoleh:

Cluster	Tahun-tahun dalam Klaster	Ciri-ciri Cluster
1	1976	jumlah penduduk miskin sangat besar, kondisi ekonomi lemah.
2	1980, 1981, 1984, 1988	Fase peralihan, mulai terlihat penurunan kemiskinan secara bertahap, dampak pembangunan awal.
3	1987, 1990, 1990*, 1993, 1993*	Tahap perbaikan dan reformasi data, penurunan kemiskinan signifikan, kebijakan lebih efektif.

Visualisasi hasil klaster melalui PCA menunjukkan pemisahan yang cukup jelas antar kelompok, masing-masing merepresentasikan fase penting dalam perjalanan pengentasan kemiskinan di Indonesia:

- a. Klaster 1 mencerminkan fase kemiskinan ekstrem, saat kondisi ekonomi nasional masih sangat terbatas.
- b. Klaster 2 menunjukkan periode transisi, ketika dampak pembangunan mulai terasa namun belum sepenuhnya merata.
- c. Klaster 3 menggambarkan tahap pemulihan dan penguatan kebijakan

3.1.10 Hasil Analisis Cluster K-Means pada Penyakit Kronis Lansia

Data yang digunakan dalam analisis ini berasal dari survei PERGEMI (Perhimpunan Gerontologi Medik Indonesia) tahun 2022, yang mencatat prevalensi 17 kategori penyakit kronis yang umum diderita oleh lansia di Indonesia. Atribut utama yang dianalisis adalah jenis penyakit dan persentase prevalensi. Sebelum dilakukan pengelompokan, ditentukan jumlah klaster optimal menggunakan metode Elbow. Hasil visualisasi elbow menunjukkan bahwa tiga klaster merupakan jumlah yang optimal. Setelah diterapkan algoritma K-Means dengan jumlah klaster $K = 3$, diperoleh pengelompokan penyakit kronis lansia sebagai berikut:

Cluster	Jenis Penyakit dalam Klaster	Ciri-ciri Cluster
0	Hipertensi, Diabetes	Prevalensi Tinggi ($\geq 20\%$)
1	Rematik, Jantung, Asthma, Asam Lambung, Lainnya	Prevalensi Sedang (5–15%)
2	Penyakit Paru Kronis, Kolesterol, Ginjal, Tumor, Osteoporosis, Anemia, Kanker, Saraf Kejepit, Liver	Prevalensi Rendah ($< 5\%$)

Visualisasi hasil clustering menunjukkan pemisahan yang jelas antar kelompok berdasarkan Tingkat prevalensi. Penyakit dengan prevalensi tinggi mendominasi fokus utama intervensi kesehatan lansia.

3.2 Pembahasan

3.2.1 Pembahasan Bitcoin

Hasil clustering memberikan gambaran bahwa perjalanan distribusi Bitcoin terbagi ke dalam tiga fase utama:

- Cluster 1 (Early): tahun 2009-2012, saat Bitcoin masih sangat awal dan pasokan masih di bawah 20%. Pada fase ini, adopsi publik masih rendah dan volatilitas tinggi.
- Cluster 2 (Growth): tahun 2013-2020, di mana supply bertumbuh pesat mendekati 75-87%. Pada fase ini, terjadi peningkatan minat investor dan awareness publik meningkat.
- Cluster 3 (Mature): setelah tahun 2021, supply mendekati maksimum. Kelangkaan supply makin nyata, dan harga Bitcoin lebih dipengaruhi oleh demand pasar, bukan lagi supply mining.

Dengan demikian, penggunaan K-Means Clustering membantu memahami fase distribusi supply Bitcoin yang berimplikasi pada strategi investasi:

- Investor bisa menyesuaikan strategi dengan siklus distribusi Bitcoin.
- Fase mature menunjukkan semakin berkurangnya pengaruh mining terhadap harga.

3.2.2 Pembahasan Saham Citigroup

Berdasarkan hasil clustering, dapat disimpulkan bahwa pergerakan harga saham Citigroup selama periode analisis dapat dibagi menjadi tiga fase utama:

- Cluster 1 (Crisis Period 2008–2011) Harga saham berada di titik terendah, dipengaruhi oleh krisis finansial global yang menghantam sektor perbankan dan keuangan. Harga saham Citigroup turun drastis hingga di bawah \$5 per lembar.
- Cluster 2 (Recovery Period 2012–2018) Memasuki fase pemulihan setelah krisis, harga saham menunjukkan tren naik yang konsisten, meskipun masih fluktuatif. Harga berkisar antara \$30–\$60 per lembar.
- Cluster 3 (Stable/Modern Period 2019–2024) Saham Citigroup relatif stabil dengan kisaran harga lebih tinggi dan volatilitas yang lebih terkontrol. Periode ini juga dipengaruhi oleh kebijakan moneter global dan kondisi makroekonomi pasca-pandemi.

Melalui penggunaan metode K-Means Clustering, diperoleh gambaran fase-fase historis pergerakan harga saham Citigroup secara lebih objektif. Informasi ini penting bagi:

- Investor yang ingin memahami siklus harga jangka panjang.
- Analisis teknikal yang ingin menyesuaikan strategi entry dan exit berdasarkan fase pasar.

3.2.3 Pembahasan Data Pajak Daerah

Dari hasil clustering, diperoleh pemetaan daerah ke dalam tiga kelompok utama:

- Cluster 1 (Low Tax Performance) Daerah-daerah dengan % Pajak rendah dan PAD rendah. Hal ini bisa disebabkan oleh:

DOI: <https://doi.org/10.31004/riggs.v4i2.1361>

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

- a) Basis ekonomi yang masih lemah.
 - b) Potensi pajak belum optimal.
 - c) Rendahnya kepatuhan atau efektivitas pemungutan pajak.
- b. Cluster 2 (Developing Region) Daerah-daerah dengan % Pajak sedang dan PAD sedang. Ini menunjukkan adanya:
 - a) Peningkatan produktivitas daerah.
 - b) Perbaikan administrasi perpajakan.
 - c) Pertumbuhan ekonomi lokal yang mulai mendorong PAD.
 - c. Cluster 3 (High Performance Region) Daerah-daerah dengan % Pajak tinggi dan PAD tinggi. Ciri-cirinya:
 - a) Ekonomi daerah kuat.
 - b) Sistem pemungutan pajak efektif.
 - c) PAD berkontribusi signifikan dalam APBD.
 - d. Dengan metode K-Means Clustering, pemerintah daerah atau pusat dapat:
 - a) Memetakan prioritas intervensi fiskal.
 - b) Memberikan bantuan teknis pada daerah Cluster 1.
 - c) Mendorong inovasi pajak di daerah Cluster 2.

3.2.4 Pembahasan Garis Kemiskinan

Dari hasil clustering, diperoleh pemetaan tahun-tahun dalam periode 1976–1993 ke dalam tiga kelompok utama, yang mencerminkan karakteristik kondisi kemiskinan nasional saat itu:

- a. Cluster 1 (Fase Kemiskinan Tinggi – 1976), Tahun 1976 mencerminkan kondisi kemiskinan paling parah sepanjang periode yang ditinjau. Angka kemiskinan desa mencapai lebih dari 40%, dan jumlah total penduduk miskin mencapai lebih dari 54 juta jiwa.
- b. Cluster 2 (Fase Transisi – 1980, 1981, 1984, 1988), Periode ini menunjukkan tahap peralihan, di mana dampak awal dari pembangunan ekonomi mulai terlihat. Terjadi penurunan bertahap dalam jumlah dan persentase penduduk miskin, meskipun efeknya belum merata di seluruh wilayah.
- c. Cluster 3 (Fase Perbaikan dan Reformasi Data – 1987, 1990, 1990*, 1993, 1993*), Klaster ini menunjukkan tren penurunan kemiskinan yang lebih konsisten. Pada periode ini, jumlah penduduk miskin menurun drastis menjadi sekitar 25 juta jiwa, dan persentasenya menurun ke kisaran 13–16 persen.

3.2.5 Pembahasan Penyakit Kronis lansia

Berdasarkan hasil clustering, penyakit kronis pada lansia di Indonesia dapat diklasifikasikan ke dalam tiga kelompok:

- a. Cluster 0 (Prevalensi Tinggi): Hipertensi (37,8%) dan Diabetes (22,9%) termasuk dalam klaster ini. Keduanya merupakan penyakit utama yang memerlukan prioritas dalam kebijakan intervensi karena dampaknya yang luas pada kualitas hidup lansia serta risiko komplikasi serius.
- b. Cluster 1 (Prevalensi Sedang): Rematik, Jantung, Asthma, Asam Lambung, serta gabungan kategori penyakit lain (seperti Stroke, Vertigo, Hernia) memiliki tingkat prevalensi sedang. Meskipun tidak setinggi cluster 0, penyakit ini tetap signifikan karena dapat memperburuk kondisi lansia jika tidak ditangani secara tepat.
- c. Cluster 2 (Prevalensi Rendah): Termasuk penyakit seperti kanker, liver, osteoporosis, dan lainnya yang memiliki prevalensi di bawah 5%. Walau secara statistik lebih kecil, keberadaan penyakit ini tetap relevan terutama karena beberapa bersifat kronis progresif dan berpotensi menyebabkan beban ekonomi serta emosional yang tinggi.

Clustering ini membantu menyusun prioritas penanganan dan intervensi berdasarkan tingkat risiko penyakit, serta memberikan dasar kuat untuk pembuatan kebijakan kesehatan lansia berbasis data.

4. KESIMPULAN

Penelitian ini menunjukkan bahwa algoritma K-Means Clustering mampu mengelompokkan pola tersembunyi dari berbagai jenis data nyata yang berasal dari domain finansial, sosial ekonomi, hingga

kesehatan. Pada dataset Bitcoin, klasterisasi berhasil mengidentifikasi tiga fase distribusi yang mencerminkan siklus pasar. Saham Citigroup dikelompokkan ke dalam tiga fase historis yang merefleksikan kondisi ekonomi pasca krisis global. Data pajak daerah berhasil memetakan wilayah berdasarkan performa fiskal, sedangkan data kemiskinan menunjukkan transformasi kesejahteraan masyarakat Indonesia secara temporal. Klasterisasi pada data penyakit kronis lansia mengungkap segmentasi risiko prevalensi yang penting untuk pengambilan kebijakan kesehatan. Secara keseluruhan, hasil ini menegaskan fleksibilitas dan efektivitas algoritma K-Means dalam mengekstraksi struktur tersembunyi pada data multidomain. Temuan dari setiap dataset memberikan wawasan yang aplikatif untuk penyusunan strategi berbasis data, baik di bidang investasi, kebijakan fiskal, maupun pelayanan kesehatan publik.

Referensi

1. Han, J., Kamber, M., & Pei, J. (2012). *Data mining: Concepts and techniques* (3rd ed.). Morgan Kaufmann. <https://archive.org/details/the-morgan-kaufmann-series-in-data-management-systems-jiawei-han-micheline-kambe>
2. Investopedia. (n.d.). *Bitcoin supply data & halving info*. <https://www.investopedia.com/terms/b/bitcoin-halving.asp>
3. Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, 31(8), 651–666. <https://doi.org/10.1016/j.patrec.2009.09.011>
4. Nasution, F. N., Lubis, M. A. R., & Purba, R. (2021). Penerapan algoritma K-Means clustering dalam analisis penyakit di Puskesmas Kota Medan. *Jurnal Mantik Penusa*, 5(2), 673–680. <https://ejournal.stmik-budidarma.ac.id/index.php/mantik/article/view/2337>
5. PERGEMI. (2022). *Survei kondisi kesehatan dan kesejahteraan lansia di Indonesia tahun 2022*. Jakarta: Perhimpunan Gerontologi Medik Indonesia. <https://www.pergemi.id/info/5/survei-kondisi-kesehatan-dan-kesejahteraan-lansia-di-indonesia>
6. Putra, R. A., & Nugroho, H. (2023). Analisis klaster harga cryptocurrency menggunakan K-Means clustering. *Jurnal Teknologi dan Sistem Informasi*, 4(1), 55–63. <https://doi.org/10.21070/jtsi.v4i1.1697>
7. Scikit-learn. (n.d.). *KMeans - sklearn.cluster.KMeans documentation*. <https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html>
8. Soesastro, H. (1995). *Mengukur kemiskinan di Indonesia: Data, masalah, dan solusi*. Jakarta: BPS & LP3ES. <https://opac.perpusnas.go.id/DetailOpac.aspx?id=135184>
9. Wulandari, R., & Iryanie, S. (2020). *Ekonomi daerah dan pajak daerah: Pendekatan teoritis dan empiris*. Jakarta: Salemba Empat. <https://opac.perpusnas.go.id/DetailOpac.aspx?id=1303772>
10. Yahoo Finance. (n.d.). *Citigroup Inc. stock historical data*. <https://finance.yahoo.com/quote/C/history>