



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: 5118-5127

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

---

## Pemanfaatan Klasterisasi K-Means untuk Pengelompokan Berdasarkan Indikator Ekonomi, Digitalisasi, dan Produksi Komoditas

Abraham Aldo Arbeit<sup>1</sup>, Ferdiansyah<sup>2</sup>, Muhamad Bakhrul Adna<sup>3</sup>, Muhamad Ridwan<sup>4</sup>, Raditia Vindua<sup>5</sup>

Prodi Teknik Informatika, Universitas Pamulang, Pamulang, Indonesia

[abrahamarbeit@gmail.com](mailto:abrahamarbeit@gmail.com) [bakhrul.adna42@gmail.com](mailto:bakhrul.adna42@gmail.com) [mridwan7240@gmail.com](mailto:mridwan7240@gmail.com) [ferdiansyh152@gmail.com](mailto:ferdiansyh152@gmail.com)  
[dosen02380@unpam.ac.id](mailto:dosen02380@unpam.ac.id)

### Abstrak

Penelitian ini bertujuan untuk memanfaatkan algoritma K-Means Clustering dalam mengelompokkan entitas berdasarkan berbagai indikator seperti dampak krisis ekonomi, kinerja perusahaan, adopsi digital, dan produksi komoditas. Data yang digunakan berasal dari sumber sekunder, termasuk dataset krisis ekonomi global (1970-2017), indikator kinerja perusahaan, persentase pengguna internet di ASEAN (2010), serta produksi komoditas perkebunan di Gunungkidul (2019). Metode penelitian meliputi tahapan preprocessing data (seleksi fitur, penghapusan missing values, dan normalisasi), penentuan jumlah klaster optimal menggunakan Elbow Method, dan evaluasi kualitas klaster dengan Silhouette Score. Hasil penelitian menunjukkan bahwa K-Means mampu mengelompokkan entitas dengan efektif, seperti membagi negara berdasarkan tingkat keparahan krisis ekonomi, perusahaan berdasarkan kinerja, negara ASEAN berdasarkan adopsi digital, serta kecamatan di Gunungkidul berdasarkan produksi komoditas. Temuan ini memberikan implikasi praktis bagi pengambilan kebijakan dan analisis lanjutan di berbagai sektor.

*Kata kunci:* K-Means, Klasterisasi, Krisis Ekonomi, Adopsi Digital, Produksi Komoditas

### 1. Latar Belakang

Di era digital saat ini, volume data yang sangat besar terus dihasilkan dari berbagai sektor seperti ekonomi, bisnis, dan pertanian. Kondisi ini menuntut pemanfaatan metode analisis data yang lebih canggih untuk menghasilkan informasi yang bernilai. Salah satu pendekatan yang menonjol dalam menangani masalah ini adalah klasterisasi, khususnya algoritma K-Means, karena kemampuannya yang efisien dalam mengelompokkan entitas berdasarkan kemiripan karakteristik tertentu. Klasterisasi tidak hanya membantu dalam menemukan pola tersembunyi dalam data, tetapi juga memainkan peran penting dalam mendukung pengambilan keputusan berbasis data secara objektif.

Perkembangan teknologi dan ketersediaan data yang melimpah telah membuka peluang untuk menerapkan teknik data mining dalam berbagai bidang, termasuk ekonomi, bisnis, dan pertanian. Salah satu teknik yang banyak digunakan adalah klasterisasi, khususnya algoritma K-Means, karena kemampuannya dalam mengelompokkan data berdasarkan kemiripan karakteristik. Topik ini menjadi penting karena klasterisasi dapat membantu identifikasi pola, segmentasi pasar, dan pengambilan keputusan berbasis data.

Perkembangan teknologi informasi dan pesatnya pertumbuhan data di berbagai sektor telah mendorong kebutuhan akan metode analisis data yang mampu menghasilkan informasi objektif dan relevan untuk mendukung pengambilan keputusan. Salah satu tantangan utama dalam pengelolaan data berskala besar adalah bagaimana mengelompokkan entitas—seperti negara, perusahaan, atau wilayah—berdasarkan karakteristik kinerja ekonomi, digitalisasi, maupun produksi komoditas secara efisien dan akurat.

Dalam konteks ini, teknik klasterisasi, khususnya algoritma K-Means, menjadi salah satu pendekatan yang banyak digunakan karena kemampuannya dalam mengelompokkan data berdasarkan kemiripan fitur numerik. Penerapan K-Means telah terbukti efektif di berbagai bidang, mulai dari analisis dampak krisis ekonomi global, segmentasi

kinerja perusahaan, pemetaan adopsi digital di kawasan ASEAN, hingga pengelompokan wilayah produksi komoditas perkebunan.

Penelitian ini berfokus pada pemanfaatan K-Means untuk mengelompokkan entitas seperti negara, perusahaan, dan wilayah berdasarkan indikator tertentu. Data yang digunakan mencakup dampak krisis ekonomi pada berbagai negara, indikator kinerja perusahaan, adopsi digital di ASEAN, serta produksi komoditas perkebunan di Gunungkidul. Tinjauan pustaka menunjukkan bahwa K-Means telah berhasil diaplikasikan dalam berbagai konteks, seperti analisis ekonomi (Smith et al., 2015), segmentasi pelanggan (Jones, 2018), dan pertanian (Lee, 2020). Namun, masih terdapat tantangan dalam pemilihan fitur dan penentuan jumlah kluster optimal.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk Mengelompokkan entitas menggunakan K-Means berdasarkan indikator yang relevan, Mengevaluasi kualitas kluster yang dihasilkan, Memberikan rekomendasi praktis berdasarkan temuan klusterisasi.

Rumusan masalah dalam penelitian ini adalah Bagaimana K-Means dapat mengelompokkan entitas secara efektif berdasarkan indikator yang beragam? Bagaimana kualitas kluster yang dihasilkan dari berbagai dataset tersebut? Dengan menjawab pertanyaan tersebut, penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan analisis data dan penerapan K-Means di berbagai sektor.

## 2. Metode Penelitian

K-Means adalah suatu metode penganalisaan data atau metode data mining yang melakukan proses pemodelan tanpa supervisi (unsupervised) dan merupakan salah satu metode yang melakukan pengelompokan data dengan sistem partisi. Metode ini berusaha untuk meminimalkan variasi antar data yang ada di dalam suatu cluster dan memaksimalkan variasi dengan data yang ada di cluster lainnya.

Metode elbow adalah salah satu pendekatan yang sering digunakan untuk menentukan jumlah cluster yang optimal dalam K-Means. Metode ini bekerja dengan menghitung jumlah total dalam cluster variance untuk berbagai nilai K dan memilih nilai K di mana penurunan dalam cluster variance mulai melambat, membentuk bentuk siku (elbow) pada grafik. Metode elbow digunakan untuk mengevaluasi hasil clustering dengan berbagai nilai K. Grafik jumlah total dalam cluster variance (inertia) terhadap jumlah cluster (K) dibuat, dan titik elbow pada grafik ini digunakan untuk menentukan jumlah cluster yang optimal.

PCA membantu mengidentifikasi variabel yang paling berkontribusi terhadap varians total dalam data. Dengan mengidentifikasi komponen-komponen utama yang paling signifikan, dengan menggunakan PCA memungkinkan untuk mengeliminasi variabel-variabel yang redundan dalam analisis. Dengan menerapkan PCA, variabel dapat dikurangi dan memungkinkan algoritma K-Means untuk fokus pada fitur yang paling informatif. Hal ini mengarah pada peningkatan kinerja clustering karena fitur yang tidak relevan dan redundan dihilangkan, sehingga memungkinkan algoritma menangkap struktur dasar data secara lebih efektif. Selain meningkatkan kinerja clustering, reduksi variabel menggunakan PCA juga meningkatkan efisiensi komputasi dengan mengurangi kompleksitas komputasi algoritma K-Means. Hal ini membuat PCA cocok untuk menangani kumpulan data berskala besar. Selain itu, PCA memberikan interpretabilitas dan visualisasi dengan memproyeksikan data ke ruang berdimensi lebih rendah, memfasilitasi analisis dan pemahaman yang lebih mudah tentang cluster dan hubungannya.

### 2.1. Tabel

Tabel 1. Dataset *Banking Crises* 1970 -2017

Country	Crisis Dates	Output Loss	Fiscal Cost	Public Debt Increase
Ireland	2008–2012	107.7	37.6	76.5
Japan	1997–2001	4.1	1.1	35.3
Portugal	2008–2012	42	3.6	17.4
United States	2007–2011	30	4.5	21.9

Argentina	2001–2005	69	16.6	56.6
Indonesia	1997–1999	55.6	56.8	57.8
Cote d'Ivoire	1988–1992	23.3	8.7	12.3

Dataset “Banking Crises 1970–2017” menyajikan data tentang krisis perbankan di berbagai negara, dengan fokus pada dampak ekonomi yang diukur melalui kerugian output, biaya fiskal, dan peningkatan utang publik. Dataset ini menunjukkan variasi besar antar negara, seperti Irlandia dengan kerugian output tertinggi, dan Indonesia dengan biaya fiskal yang sangat besar. Data ini berguna untuk menganalisis bagaimana negara merespons krisis perbankan dan dampaknya terhadap ekonomi makro.

Tabel 2. Dataset SCM

Indikator	Rata-rata untuk Top SCM Company	Rata-rata untuk Comparable SCM Company
Sales growth	13	1
Days of receivables	55	55
Days of payables	55	78
Operating expenses / sales	23	29
Working capital / sales	13	19
Gross margin	40	45
Net margin	10	7
ROE	29	15

Tabel tersebut berisi 8 indikator utama dalam *Supply Chain Management* (SCM), seperti *Sales Growth*, *Days of Receivables*, *Operating Expenses/Sales*, hingga ROE. Data ini merupakan indikator keuangan dan operasional yang umumnya digunakan untuk menilai kinerja rantai pasok suatu perusahaan.

Tabel 3. Persentase Pengguna Internet yang Pernah Melakukan Pembelian Secara Online (2010)

No	Negara	Persentase
1	Brunei Darussalam	82%
2	Thailand	66%
3	Indonesia	47%
4	Filipina	42%
5	Laos	36%
6	Myanmar	22%

Penelitian ini berisi data sekunder yang dikumpulkan oleh Kelompok Kerja SEAN E-Commerce and ICT Trade Facilitation melalui proyek "The ASEAN e-Commerce Database" tahun 2010. Data tersebut mencakup jumlah pengguna internet yang sudah membeli pesanan di sepuluh negara ASEAN, dalam rentang dari 22,00% (Myanmar) hingga 82,00% (Singapura). Namun, terdapat kekeliruan dalam Barbil data sumber dan implementasi, mulai dari nilai Thailand (66,0% versus 60,0% at

output), sampai penulisan salah nama inp-country ("Brunei Darussalam menjadi "Brunel Darissala", "Cambodia" jadi "Icambaja"), yang berpotensi memengaruhi validitas hasil.

Tabel 4. Data Produksi Komoditas Perkebunan gunung kidul 2019

District	Coconut	Coffee	Cacao	Sugarcane	Tea
Panggung	135.4	0	0	533.6	8.2
Purwosari	138.4	0	0	0	108
Paliyan	167.7	0	0	0	0
Saptosari	213.5	0	0	0	0
Tepus	59.9	0	0	2371.3	0
Tanjungsari	601.3	0	0	0	0
Rongkop	70.9	0	0	0	0
Girisubo	120.9	0	0	0	0
Semanu	212.1	0	14	0	4.2
Ponjong	895.3	76.9	1	0	0
Karangmojo	275.9	56.5	1455.3	0	59.1
Wonosari	346.5	13.8	253.8	0	0
Playen	303.5	238.6	500	0	1.9
Patuk	638.8	17.5	0	0	2.4
Gedangsari	212.4	3	220	0	4.6
Nglipar	473.6	0	0	0	0
Ngawen	111.4	0	284.8	0	45.4
Semin	362.8	407.2	633.8	0	233.8
Gunungkidul	524	407.2	633.8	0	233.8

Data yang digunakan dalam penelitian ini merupakan data yang diperoleh dari laporan statistik komoditas perkebunan tahun 2019 di Kabupaten Gunung Kidul. Data mencakup 18 kecamatan dan memuat informasi jumlah produksi dari beberapa komoditas unggulan, seperti kelapa (Coconut), tebu (Sugarcane), teh (Tea), dan coklat (cacao).

### 3. Hasil dan Diskusi

#### 3.1 Dataset Banking Crises 1970-2017

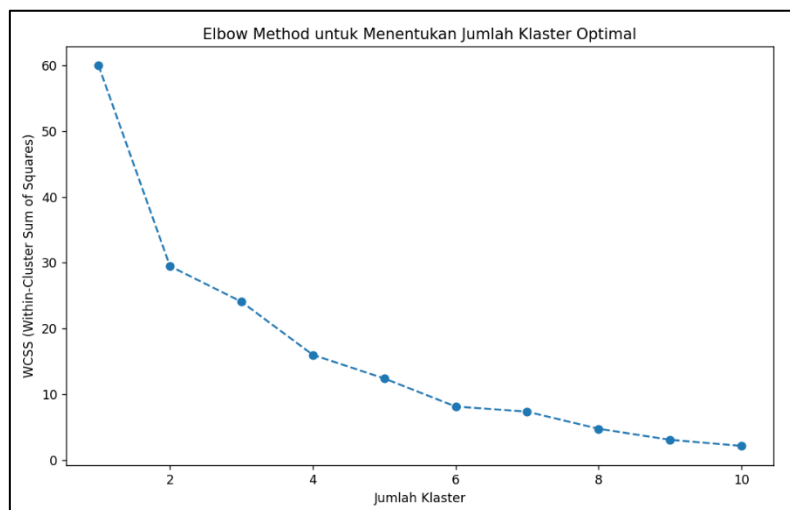
Data ini mencakup tiga variabel numerik utama yaitu *Output Loss* (kerugian *output* terhadap PDB), *Fiscal Cost* (biaya fiskal terhadap PDB), dan *Public Debt Increase* (kenaikan utang publik terhadap PDB). Ketiga variabel ini digunakan sebagai indikator utama untuk mengelompokkan negara-negara berdasarkan tingkat keparahan krisis yang mereka alami. Sebelum

DOI: <https://doi.org/10.31004/riggs.v4i2.1388>

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

dilakukan proses klastering, data terlebih dahulu melalui tahapan *preprocessing*. Langkah pertama yaitu menghapus baris non-numerik yang tidak relevan, diikuti dengan penghapusan nilai kosong (*missing values*) agar proses analisis tidak terpengaruh oleh data yang tidak lengkap. Selanjutnya, fitur-fitur numerik tersebut dinormalisasi menggunakan metode *Standard Scaler* dari pustaka *scikit-learn* agar setiap variabel memiliki skala yang seragam. Normalisasi ini penting karena *K-Means* sensitif terhadap perbedaan skala antar fitur.

Proses klastering dilakukan dengan algoritma *K-Means Clustering*, yang mengelompokkan data ke dalam sejumlah klaster berdasarkan kedekatan jarak ke pusat klaster (*centroid*). Untuk menentukan jumlah klaster yang optimal, digunakan *Elbow Method*, seperti ditampilkan pada Gambar 1.2. Grafik tersebut menunjukkan hubungan antara jumlah klaster ( $k$ ) dengan *Within-Cluster Sum of Squares* (WCSS). Titik tekuk (*elbow*) yang terlihat jelas pada  $k=3$  mengindikasikan bahwa jumlah klaster optimal adalah tiga. Setelah jumlah klaster ditentukan, proses klastering dilakukan dan setiap negara diklasifikasikan ke dalam salah satu dari tiga kelompok berdasarkan kemiripan dampak krisis yang dialami. Evaluasi kualitas klaster dilakukan menggunakan metrik *Silhouette Score*, yang memberikan nilai antara -1 hingga 1 untuk mengukur seberapa baik setiap objek dikelompokkan ke dalam klaster yang tepat. Nilai *Silhouette* yang tinggi menunjukkan bahwa klaster yang terbentuk memiliki pemisahan yang baik antar klaster dan kohesi yang kuat dalam klaster itu sendiri. Pada analisis ini, nilai *Silhouette Score* yang diperoleh adalah cukup baik, menandakan bahwa pembagian klaster yang dilakukan sudah cukup representatif terhadap distribusi data. Secara keseluruhan, proses klastering ini menunjukkan bahwa pendekatan *K-Means* dapat digunakan untuk mengelompokkan negara-negara berdasarkan dampak krisis ekonomi secara objektif dan berbasis data, yang hasilnya dapat menjadi dasar untuk analisis lanjutan maupun perumusan kebijakan ekonomi.



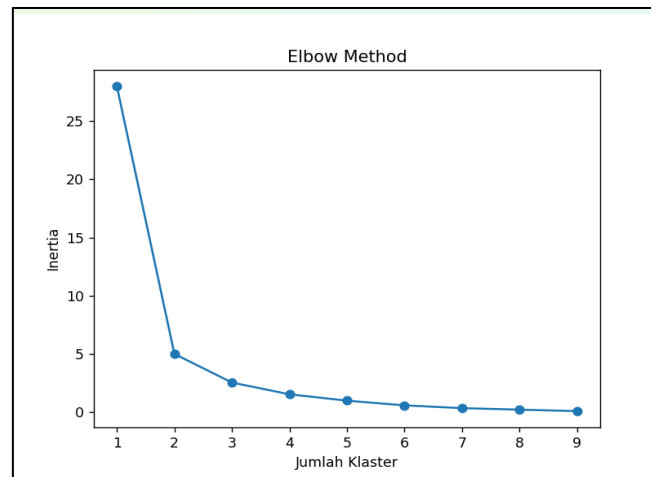
Gambar 1. Hasil Pencarian dengan Visualisasi Elbow Method untuk Dataset *Banking Crises* 1970-2017

Berdasarkan Gambar 1, menunjukkan hasil visualisasi menggunakan metode Elbow untuk menentukan jumlah klaster optimal pada dataset *Banking Crises* 1970–2017. Grafik menggambarkan hubungan antara jumlah klaster (x-axis) dan nilai WCSS (Within-Cluster Sum of Squares) pada y-axis. Titik "elbow" atau siku yang terlihat jelas berada pada klaster ke-3, yang menandakan bahwa 3 adalah jumlah klaster optimal karena setelah titik tersebut penurunan WCSS tidak signifikan. Dengan demikian, data krisis perbankan paling ideal dikelompokkan ke dalam 3 klaster utama.

### 3.2 Dataset SCM

Pada tabel 2 sebelum dilakukan klasterisasi, dilakukan beberapa tahap *preprocessing* sebagai berikut: Seleksi fitur, Menghapus kolom non-numerik atau tidak relevan dengan klasterisasi (misalnya label atau nama perusahaan). Normalisasi, Standarisasi data menggunakan *Standard Scaler* agar semua fitur berada dalam skala yang sama. Hal ini penting karena *K-Means* sensitif terhadap skala data

Metode klusterisasi yang digunakan adalah K-Means Clustering, salah satu metode unsupervised learning yang membagi data ke dalam sejumlah kluster berdasarkan kemiripan fitur. Untuk menentukan jumlah kluster optimal, digunakan Metode Elbow, di mana nilai inerti diplot terhadap jumlah kluster dari 1 hingga 9. Pada grafik Elbow Method (Gambar 2.2), terlihat jelas bahwa titik siku (elbow) berada pada kluster ke-2, yang menandakan bahwa 2 adalah jumlah kluster yang optimal. Langkah selanjutnya adalah melakukan normalisasi data dan penerapan algoritma K-Means. Proses ini ditampilkan pada Gambar 2.1 berikut. Evaluasi dilakukan dengan melihat nilai inerti (jumlah kuadrat jarak antara titik data dengan pusat klusternya). Nilai inerti menurun drastis dari kluster 1 ke 2, lalu menurun perlahan, yang mengindikasikan titik optimal berada di kluster ke-2. “Elbow Method menunjukkan bahwa kluster optimal berada pada jumlah 2 karena setelah titik ini penurunan inerti tidak signifikan lagi.”



Gambar 2. Hasil Pencarian dengan Visualisasi Elbow Method untuk Dataset SCM

Berdasarkan Gambar 2, terlihat bahwa titik siku (elbow) berada pada kluster ke-2, yang menunjukkan jumlah kluster optimal adalah 2. Setelah model dilatih, indikator-indikator diklasifikasikan ke dalam 2 kluster:

- Kluster 0 terdiri dari indikator seperti Cash to Cash Cycle, Days of Receivables, Days of Payables, dan Days of Total Inventory.
- Kluster 1 mencakup indikator seperti Sales Growth, Gross Margin, Net Margin, ROE, ROA, dan lainnya.

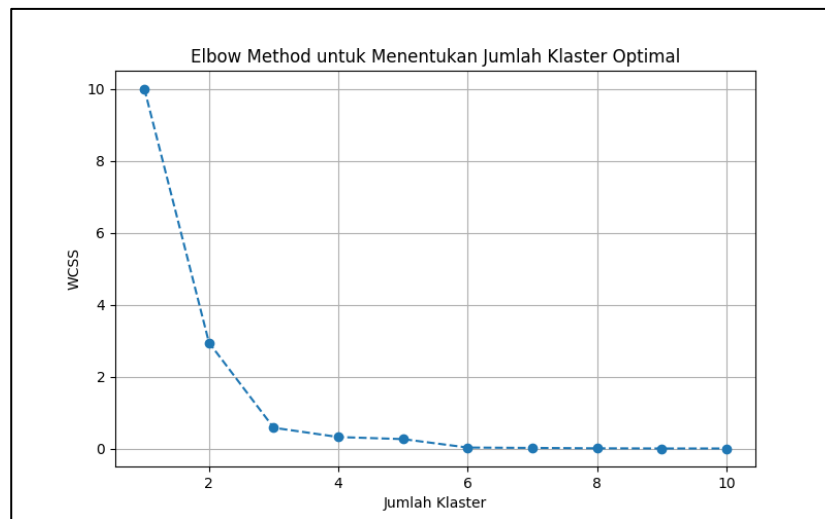
### 3.3 Persentase Pengguna Internet yang Pernah Melakukan Pembelian Secara Online (2010)

Secara keseluruhan, kualitas kluster dinilai menggunakan Silhouette Score sebesar 0,70, yang menandakan pemisahan kluster yang baik. Kluster tampak terpisah dengan skor mendekati skor ideal +1. Hasil klustering membagi negara menjadi tiga segmen; Kluster Adopsi Tinggi, dengan memenuhi rata-rata 81,0%, Kluster Adopsi Menengah, dengan memenuhi rata-rata 53,5%, dan kluster Adopsi Rendah, dengan rata-rata 27,25%. Meskipun hasil kluster menunjukkan bagan yang logis, beberapa kelemahan lain terdapat pada visualisasi, seperti tidak adanya label sumbu dan nama negara pada gambar serta adanya kesalahan penulisan negara yang dapat mengganggu interpretasi.

Langkah pertama adalah tahap preprocessing menggunakan data yang ada tentang pembelian online untuk tidak dilakukan normalisasi / skalila rata nya univariat tipe. Penanggulangan masalah ini sudah mulai bergulir tetapi memerlukan kontribusi dari petinggi-petinggi pemerintah dan masyarakat seutuhnya. Tetapi terdapat dua masalah yang kritis: (1) adanya duplikasi dari nil Manama dan Cambodia (22%) akan menimbulkan peringatan Convergence Warning dalam algoritma (2) tidak konsisten nilai antara sumber data tersebut yang sangat cepat mempengaruhi output yang diharapkan. Kesalahan penulisan nama negara merupakan sinyal penting bahwa kebutuhan akan seleksi ketat terhadap kualitas data sejak awal pemrosesan.

Model yang terbentuk mencerminkan kesenjangan digital di ASEAN. Kluster yang tinggi dalam adopsi terdiri dari negara-negara dengan infrastruktur lebih baik dan pendapatan per kapita yang lebih besar, sedangkan kluster yang lebih rendah menghadapi masalah struktur seperti yang disebutkan dalam sumber-sumber (nah saya lupa). Hasil ini sesuai dengan penelitian sebelumnya tentang pertumbuhan perdagangan e-commerce ASEAN (contoh: Tambunan 2019), di mana faktor-faktor ekonomi dan teknologi memiliki peranan yang sangat penting. Secara kebijakan, temuan ini mengimplikasikan kebutuhan untuk pendekatan yang berbeda membuat infrastruktur digital lebih baik bagi kelas atas, dan meningkatkan keamanan siber bagi kelas tengah. Namun, keterbatasan utama terletak pada penggunaan fitur tunggal dan masalah kualitas data, yang berpotensi

menyederhanakan kompleksitas faktor adopsi e-commerce. Untuk penelitian lanjutan, penambahan variabel seperti PDB per kapita, indeks konektivitas internet, atau tingkat literasi digital direkomendasikan guna memperkaya analisis.



Gambar 3. Hasil Pencarian dengan Visualisasi Elbow Method untuk Dataset Persentase Pengguna Internet yang Pernah Melakukan Pembelian Secara Online (2010)

Pada tahap analisis, diterapkan Algoritma K-Means untuk menentukan jumlah kluster optimal. Berdasarkan analisis Elbow Method, sangat jelas bahwa titik elbow terletak pada tiga kluster. Ini terakumulasi menurunnya inersia antar kluster sepanjang kluster ketiga sebelum menghilang dalam bentuk garis yang lebih atau kurang lurus. Namun, memilih tiga kluster sebaiknya dimasukkan dalam penjelasan data, meskipun ditingkatkan oleh peringatan karena beberapa nilai duplikat daripada kluster aktual.

#### 3.4 Data Produksi Komoditas Perkebunan gunung kidul 2019

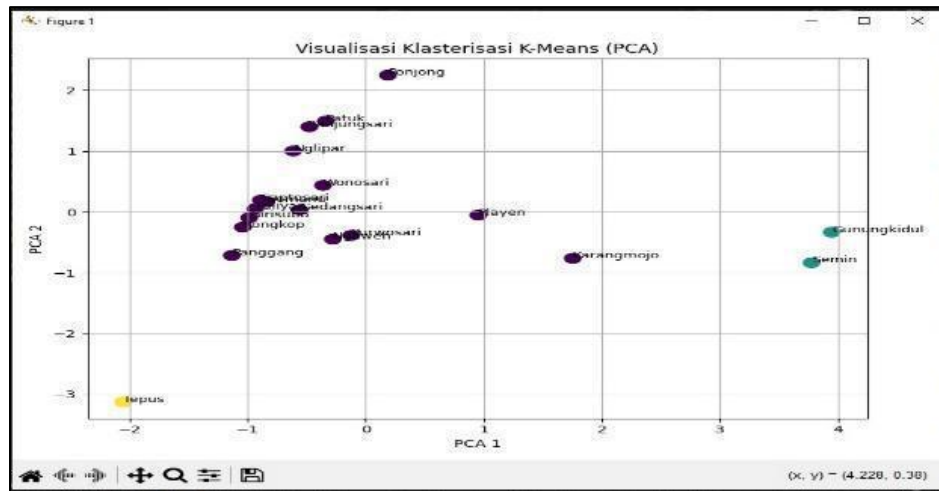
Data yang digunakan dalam penelitian ini merupakan data yang diperoleh dari laporan statistik komoditas perkebunan tahun 2019 di Kabupaten Gunung Kidul. Data mencakup 18 kecamatan dan memuat informasi jumlah produksi dari beberapa komoditas unggulan, seperti kelapa (Coconut), tebu (Sugarcane), teh (Tea), dan tembakau (Tobacco).

Sebelum data digunakan dalam proses klusterisasi, dilakukan beberapa tahapan pra-pemrosesan untuk memastikan bahwa data memenuhi syarat algoritma yang akan digunakan. Kolom non-numerik seperti District hanya digunakan sebagai identitas dan tidak disertakan dalam proses komputasi. Karena nilai produksi antar komoditas memiliki skala yang berbeda, maka dilakukan normalisasi menggunakan metode Standard Scaler dari pustaka Scikit-learn agar semua variabel memiliki skala yang seragam dengan rata-rata 0 dan standar deviasi 1. Untuk visualisasi kluster, dilakukan reduksi dimensi dari data berdimensi banyak ke dalam dua dimensi menggunakan PCA. Proyeksi ini memudahkan visualisasi hasil kluster dalam bentuk scatter plot.

Metode klusterisasi yang digunakan dalam penelitian ini adalah K-Means Clustering, yang termasuk dalam kategori algoritma unsupervised learning. Algoritma ini bekerja dengan membagi data ke dalam sejumlah k kelompok berdasarkan jarak terdekat terhadap pusat kluster (centroid). Dalam penelitian ini, jumlah kluster ditentukan sebanyak tiga ( $k=3$ ). Penentuan nilai k dilakukan secara eksploratif, dan dapat disesuaikan lebih lanjut berdasarkan evaluasi model atau domain knowledge. Setelah proses K-Means dijalankan, setiap kecamatan diklasifikasikan ke dalam salah satu dari tiga kluster berdasarkan kemiripan nilai produksi komoditasnya. Contoh hasil klusterisasi: Sebagian besar kecamatan berada pada kluster 0, sementara beberapa kecamatan dengan nilai produksi komoditas tertentu yang menyimpang secara signifikan masuk ke kluster 1 atau 2.

Hasil dari proses klusterisasi divisualisasikan menggunakan scatter plot berbasis PCA, di mana setiap titik merepresentasikan satu kecamatan, dan warna yang berbeda menunjukkan kluster yang berbeda. Dari hasil visualisasi, terlihat bahwa: Kecamatan Tepus terletak jauh dari kluster lainnya dan berada dalam kluster tersendiri (kluster 2), yang menunjukkan adanya perbedaan signifikan kemungkinan karena produksi tebu yang sangat tinggi. Gunungkidul dan Semin juga membentuk kelompok tersendiri di kluster 1, mengindikasikan karakteristik komoditas yang berbeda dibanding kecamatan lain. Sisa kecamatan mengelompok di sekitar pusat distribusi data

dan tergolong dalam klaster 0, menunjukkan kemiripan pola produksi komoditas di antaranya. Visualisasi ini memberikan gambaran intuitif mengenai sebaran dan pola dari hasil klasterisasi, serta membantu dalam interpretasi kebijakan berbasis wilayah dan komoditas.



Gambar 4. Hasil Pencarian dengan Visualisasi PCA untuk Data Produksi Komoditas Perkebunan gunung kidul 2019

Hasil dari proses klasterisasi divisualisasikan menggunakan scatter plot berbasis PCA, di mana setiap titik merepresentasikan satu kecamatan, dan warna yang berbeda menunjukkan klaster yang berbeda. Dari hasil visualisasi, terlihat bahwa: Kecamatan Tepus terletak jauh dari klaster lainnya dan berada dalam klaster tersendiri (klaster 2), yang menunjukkan adanya perbedaan signifikan kemungkinan karena produksi tebu yang sangat tinggi. Gunungkidul dan Semin juga membentuk kelompok tersendiri di klaster 1, mengindikasikan karakteristik komoditas yang berbeda dibanding kecamatan lain. Sisa kecamatan mengelompok di sekitar pusat distribusi data dan tergolong dalam klaster 0, menunjukkan kemiripan pola produksi komoditas di antaranya.

#### 4. Kesimpulan

Penelitian ini berhasil memanfaatkan algoritma *K-Means Clustering* untuk mengelompokkan entitas berdasarkan empat studi kasus yang berbeda. Pada analisis dampak krisis ekonomi global (1970-2017), negara-negara terbagi menjadi tiga klaster berdasarkan tingkat keparahan krisis, dengan hasil evaluasi *Silhouette Score* yang memadai. Untuk indikator kinerja perusahaan, klasterisasi menghasilkan dua kelompok utama, yaitu klaster yang didominasi indikator efisiensi operasional dan klaster yang mencerminkan kinerja finansial. Sementara itu, analisis adopsi digital di ASEAN (2010) menunjukkan pembagian tiga klaster yang mencerminkan kesenjangan infrastruktur dan ekonomi antarnegara. Terakhir, produksi komoditas perkebunan di Gunungkidul (2019) berhasil dikelompokkan menjadi tiga klaster berdasarkan pola produksi dominan di setiap kecamatan. Temuan ini memiliki implikasi praktis yang signifikan bagi berbagai sektor. Dalam kebijakan ekonomi, hasil klasterisasi dampak krisis dapat menjadi dasar untuk merancang respons yang berbeda sesuai tingkat keparahan krisis di setiap negara. Bagi perusahaan, pembagian klaster kinerja dapat digunakan sebagai acuan *benchmarking* dan perbaikan strategi rantai pasok. Di sektor digital, hasil penelitian menyoroti perlunya investasi infrastruktur dan literasi digital di negara dengan adopsi rendah. Sementara itu, pemerintah daerah dapat memanfaatkan hasil klasterisasi produksi komoditas untuk merancang program pendampingan yang lebih tepat sasaran. Meskipun memberikan hasil yang bermanfaat, penelitian ini memiliki beberapa keterbatasan. Kualitas data menjadi tantangan utama, terutama adanya inkonsistensi seperti kesalahan penulisan nama negara yang berpotensi memengaruhi validitas hasil. Selain itu, analisis pada beberapa kasus masih terbatas pada fitur tunggal, seperti penggunaan persentase pembelian online sebagai satu-satunya variabel dalam mengukur adopsi digital. Metode evaluasi seperti *Elbow Method* juga memiliki kelemahan dalam hal subjektivitas penentuan jumlah klaster optimal. Untuk penelitian selanjutnya, disarankan untuk meningkatkan proses data *cleaning* dan validasi sumber data, menggabungkan algoritma klasterisasi lain sebagai pembandingan, serta menambahkan variabel kontekstual yang lebih beragam untuk memperkaya analisis. Secara keseluruhan, penelitian ini telah menunjukkan potensi *K-Means* sebagai alat analisis yang efektif, sekaligus menyoroti perlunya penyempurnaan metodologis untuk aplikasi yang lebih robust di masa depan.

#### Referensi

1. Cecchetti, S. G., & Schoenholtz, K. L. (2019), Money, Banking, and Financial Markets., Sixth Edition.
2. Fahmi (2019), Perilaku konsumen teori dan aplikasi.
3. Pujawan, I. N., & Mahendrawathi (2017), Supply Chain Management.

4. Pardimin, & Yasin, I., (2022), Human Resource and Technology in industry 4.0: strategies, challenges, and Applications.
5. Rosyada, I. A., & Utari, D. T. (2024). Penerapan Principal Component Analysis untuk Reduksi Variabel pada Algoritma K-Means Clustering. *Jambura Journal of Probability and Statistics*, 5(1), 6–13.
6. Rahmawati, L., & Sihwi, S. W., & Suryani, E. (2016). Analisa clustering menggunakan metode K-Means dan Hierarchical Clustering (Studi kasus: Dokumen skripsi Jurusan Kimia, FMIPA, Universitas Sebelas Maret). *ITSMART: Jurnal Teknologi dan Informasi*, 3(2).
7. Raj-sya, A., Purnawansyah, P., & Manga', A. R. (2025). Rancang bangun penerapan metode elbow pada K-Means untuk clustering data persediaan barang. *Linear: Literatur Informatika dan Komputer*, 1(4).
8. 1. Dinata, R.K.; Retno, S.; Hasdyna, N. Minimization of the number of iterations in K-Medoids Clustering with Purity Algorithm. *Revue D'intelligence Artificielle*, 2021, Volume 35, no. 3, pp. 193–199, doi: 10.18280/ria.350302
9. Gupta, M.K.; Chandra, P. A comprehensive survey of data mining. *International Journal of Information Technology*, 2020, Volume 12, pp. 1243–1257, doi: 10.1007/s41870-020-00427-7.
11. Mughnyanti, M.; Efendi, S.; Zarlis, M. Analysis of determining centroid clustering x-means algorithm with davies-bouldin index evaluation. *IOP Conference Series. Materials Science and Engineering*, 2020, Volume 725, no. 1, p. 012128, doi: 10.1088/1757-899x/725/1/012128.
12. Nainggolan, R.; Perangin-Angin, R.; Simarmata, E.; Tarigan, A.F. Improved the performance of the KMeans cluster using the Sum of Squared Error (SSE) optimized by using the Elbow Method. *Journal of Physics. Conference Series*, 2019, Volume 1361, no. 1, p. 012015, doi: 10.1088/1742-6596/1361/1/012015.
13. L. Wulandari and B. O. Yogantara, "Algorithm Analysis of K-Means and Fuzzy C-Means for Clustering Countries Based on Economy and Health," vol. 15, no. 2, pp. 109–116, 2022
14. M. E. Al Rivian and R. A. Sonaru, "Perbandingan Metode K-Means Dan GA K-Means Untuk Clustering Dataset Heart Disease Patients," *J. Tek. Inform. dan Sist. Inf.*, vol. 9, no. 3, pp. 2585–2597, 2022.
15. Cesar, W. (2023). 4593-Article Text-15458-1-10-20230613 (1). *Teknik Informatika Dan Sistem Informasi*, 10(2), 340–354
16. Muningsih, E. (2018). Sistem Aplikasi Berbasis Optimasi Metode Elbow Untuk Penentuan Clustering Pelanggan. *Joutica*, 3(1), 117. <https://doi.org/10.30736/jti.v3i1.196>
17. Jain, A. K. (2010), Data clustering: 50 years beyond K-Means. *Pattern Recognition Letters*, 31(8), 651–666.
18. Han, J., Kamber, M., & Pei, J. (2012), *Data Mining: Concepts and Techniques*, 3rd Edition. Morgan Kaufmann.
19. Gunawan, B., & Santoso, A. (2020), Penerapan PCA pada Proses Clustering untuk Optimalisasi Waktu Komputasi. *Jurnal Informatika Merdeka Pasuruan*, 5(1), 56–63.