



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: 4479-4487

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

---

## Analisis Perbandingan Hasil K-Means Clustering untuk Sektor Ekonomi dan Kesehatan

Radithya Danutirta, Ilham Bustomi, Lisa Ardeliana, Reza Putra Nurhuda, Bagas Rudianto, Zurnan Alfian

<sup>123456</sup> Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Pamulang

\*1 [radithyadanutirta@gmail.com](mailto:radithyadanutirta@gmail.com), 2 [bustomiilham117@gmail.com](mailto:bustomiilham117@gmail.com), 3 [lissaadhelia13@gmail.com](mailto:lissaadhelia13@gmail.com), 4 [masrezaa354@gmail.com](mailto:masrezaa354@gmail.com), 5 [bagas1212@gmail.com](mailto:bagas1212@gmail.com), 6

[dosen02678@unpam.ac.id](mailto:dosen02678@unpam.ac.id)

### **Abstrak**

*Penelitian ini melakukan analisis komparatif terhadap hasil klusterisasi menggunakan algoritma K-Means Clustering pada lima dataset yang berasal dari dua sektor berbeda, yaitu ekonomi dan kesehatan. Dataset yang dianalisis meliputi data IPO perusahaan besar di Amerika Serikat, konsumsi energi dan kebutuhan oksigen kendaraan bermotor, persebaran COVID-19 di Indonesia, indeks daya beli berdasarkan harga Big Mac, serta komponen neraca pembayaran Indonesia. Setiap dataset dianalisis secara terpisah untuk mengidentifikasi pola tersembunyi dan klasifikasi alami berbasis fitur numerik. Hasil menunjukkan bahwa pada seluruh kasus, algoritma K-Means mampu mengidentifikasi struktur kluster yang jelas dan konsisten, dengan pola umum klasifikasi menjadi tiga kelompok (rendah–sedang–tinggi). Hasil ini mengindikasikan bahwa K-Means Clustering memiliki potensi sebagai alat eksplorasi data multivariabel yang efektif di berbagai domain. Temuan ini juga membuka peluang pemanfaatan klusterisasi sebagai dasar segmentasi strategis dalam perumusan kebijakan dan analisis pasar.*

*Kata kunci: K-Means Clustering, IPO, COVID-19, Neraca Pembayaran, Purchasing Power Parity, Polusi Kendaraan*

### **1. Latar Belakang**

Perkembangan teknologi informasi, khususnya dalam bidang *big data* dan kecerdasan buatan, telah merevolusi pendekatan analisis data di berbagai sektor, termasuk ekonomi dan kesehatan. Tantangan utama kini tidak lagi pada pengumpulan data, melainkan bagaimana mengekstraksi pola tersembunyi dari data yang kompleks dan berskala besar. Salah satu metode yang banyak digunakan dalam pengelompokan data numerik adalah *K-Means Clustering*, yang dikenal karena kesederhanaannya, efisiensi komputasi, serta kemampuannya dalam mengungkap struktur laten.

Sejauh ini, penerapan *K-Means* umumnya masih terbatas pada studi satu domain, seperti pengelompokan kinerja IPO perusahaan, analisis daya beli melalui indeks *Big Mac*, serta pemetaan penyebaran COVID-19. Namun, kajian yang menggabungkan sektor ekonomi dan kesehatan dalam satu kerangka analisis masih jarang ditemukan, padahal keduanya saling berkaitan erat dan berdampak langsung terhadap kebijakan publik.

Penelitian ini bertujuan mengisi celah tersebut dengan menerapkan *K-Means Clustering* pada lima dataset multidomain dari sektor ekonomi dan kesehatan. Tujuannya adalah mengeksplorasi terbentuknya kluster alami dan menilai efektivitas metode ini dalam segmentasi serta visualisasi data lintas sektor.

Cakupan analisis meliputi pengelompokan kendaraan bermotor berdasarkan konsumsi energi dan kebutuhan oksigen sebagai indikator polusi, serta pemetaan karakteristik COVID-19 antarprovinsi. Dari sisi ekonomi, mencakup analisis IPO perusahaan besar di Amerika Serikat, evaluasi nilai tukar melalui harga *Big Mac* antarnegara, dan segmentasi komponen neraca pembayaran Indonesia dari tahun 2018 hingga kuartal II 2025.

Melalui pendekatan ini, penelitian diharapkan dapat menilai potensi *K-Means* dalam mendukung pengambilan keputusan berbasis data yang integratif dan kontekstual.

## 2. Metode Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan algoritma *K-Means Clustering* untuk menganalisis dan mengelompokkan data dari berbagai domain. Implementasi dilakukan menggunakan bahasa pemrograman *Python* dengan pustaka *pandas*, *scikit-learn*, *matplotlib*, dan *seaborn*. Proses pengolahan data dan visualisasi hasil klusterisasi dilakukan dalam lingkungan kerja *Jupyter Notebook* dan *Visual Studio Code*.

Terdapat lima dataset yang digunakan dalam penelitian ini. Pertama, data IPO terbesar di pasar saham Amerika Serikat diperoleh dari buku *Finance for Executives* (2019) dan situs [www.renaissancecapital.com/IPO-Center/Stats/Largest-US-IPO](http://www.renaissancecapital.com/IPO-Center/Stats/Largest-US-IPO) dengan fokus pada dua variabel utama yaitu *Amount Raised* dan *Market Value at IPO*. Kedua, data kebutuhan oksigen kendaraan bermotor yang bersumber dari buku *Kesehatan Masyarakat Perkotaan* (2022) dan seminar Kota Sehat Universitas Hasanuddin, digunakan untuk menganalisis konsumsi energi dan dampak lingkungan kendaraan.

Selanjutnya, data ketiga berasal dari statistik COVID-19 nasional tahun 2020–2022, yang berisi informasi jumlah kasus, tingkat kesembuhan, dan tingkat kematian di 34 provinsi Indonesia. Data ini digunakan untuk mengelompokkan wilayah berdasarkan karakteristik pandemi. Dataset keempat menggunakan *Big Mac Index* yang diambil dari buku *International Financial Management* (2018) terbitan McGraw-Hill, yang digunakan untuk mengilustrasikan konsep *Purchasing Power Parity* (PPP) dengan membandingkan harga Big Mac dan nilai tukar mata uang terhadap USD di berbagai negara.

Terakhir, data kelima berasal dari buku *Statistik Keuangan dan Neraca Pembayaran Indonesia* (Penerbit Andi, 2021), yang mencakup komponen-komponen neraca pembayaran Indonesia dari tahun 2018 hingga kuartal II tahun 2025. Data ini meliputi transaksi berjalan, ekspor, impor, jasa, pendapatan primer, sekunder, dan transaksi finansial. Semua dataset tersebut dinormalisasi terlebih dahulu sebelum dilakukan proses klusterisasi menggunakan metode *K-Means*, dengan penentuan jumlah kluster optimal.

## 3. Hasil dan Diskusi

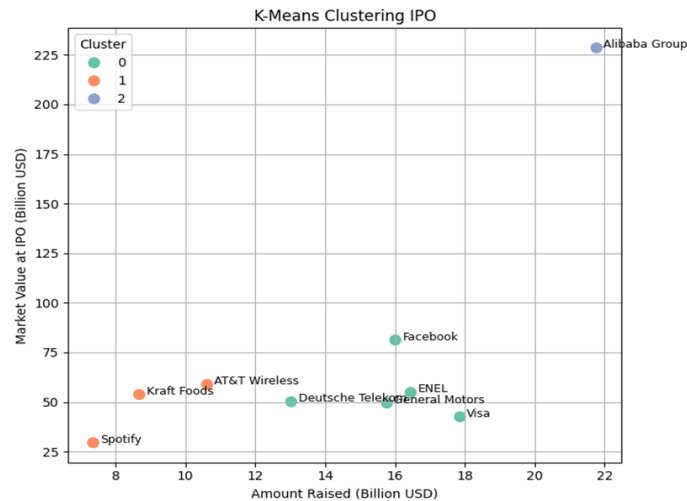
### 3.1. Klusterisasi IPO Terbesar di Amerika Serikat

Dataset ini terdiri atas data IPO dari perusahaan besar di US, dengan dua fitur utama: *Amount Raised* (dana dihimpun) dan *Market Value at IPO* (valuasi saat penawaran saham perdana). Hasil klusterisasi menunjukkan tiga kelompok utama:

- **Cluster 0 (Hijau):** IPO besar-menengah seperti *Facebook*, *Visa*, dan *ENEL*
- **Cluster 1 (Oranye):** IPO kecil seperti *Spotify* dan *Kraft Foods*
- **Cluster 2 (Biru):** IPO dengan nilai ekstrem, yaitu *Alibaba Group*

Tabel 1. Dataset IPO dari Buku Finance for Executives

Company	Country	Year	Exchange	Amount raised	Market value at IPO
Alibaba Group	China	2014	NYSE	21,77	228,53
Visa	USA	2008	NYSE	17,86	42,53
ENEL	Italy	1999	NYSE	16,45	54,85
Facebook	USA	2012	NASDAQ	16,01	81,25
General Motors	USA	2010	NYSE	15,77	49,5
Deutsche Telekom	Germany	1996	NYSE	13,03	50,11
AT&T Wireless	USA	2000	NYSE	10,62	58,85
Kraft Foods	USA	2001	NYSE	8,68	53,79
Spotify	Sweden	2018	NYSE	7,36	29,5



Gambar 1. Visualisasi K-Means Clustering pada Data IPO

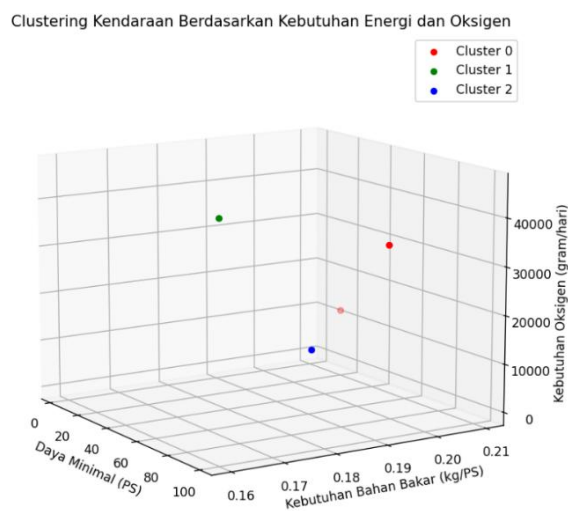
### 3.2. Kebutuhan Energi dan Oksigen Kendaraan Bermotor

Dataset ini berisi informasi daya kendaraan, konsumsi bahan bakar, dan kebutuhan oksigen harian sebagai indikator dampak polusi lingkungan. Hasil klasterisasi menunjukkan tiga tipe:

- **Cluster 0 (Merah):** Mobil dan Truk (kebutuhan sedang)
- **Cluster 1 (Hijau):** Bus (kebutuhan tinggi)
- **Cluster 2 (Biru):** Motor (kebutuhan rendah)

Tabel 2. Dataset Kendaraan dari Buku Kesehatan Masyarakat Perkotaan

Jenis Kendaraan	Daya Minimal (PS)	Kebutuhan Bahan Bakar (kg/PS)	Kebutuhan Oksigen (kg/hari)	Kebutuhan Oksigen (gram/hari)
Motor	1	0,21	0,5817	581,7
Mobil	20	0,21	11,634	11634
Truk	50	0,21	29,085	29085
Bus	100	0,16	4576	45760



Gambar 2. Visualisasi 3D Klasterisasi Kendaraan Berdasarkan Konsumsi Energi dan Oksigen

### 3.3. Klasterisasi Provinsi Berdasarkan Data *COVID-19*

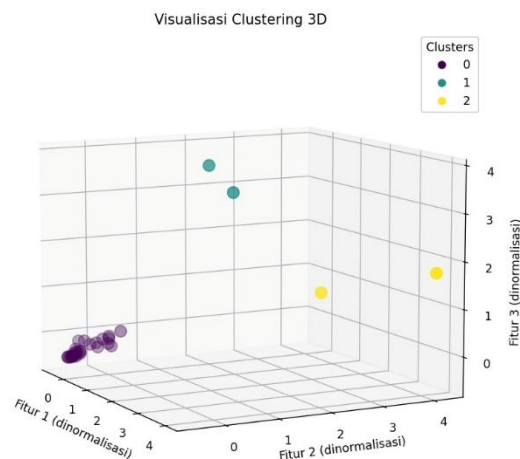
Dataset *COVID-19* mencakup 34 provinsi Indonesia dengan fitur utama: total kasus, tingkat kesembuhan, dan tingkat kematian. Hasil klasterisasi:

- **Cluster 0 (Ungu):** Mayoritas provinsi dengan karakteristik rata-rata
- **Cluster 1 (Hijau):** Provinsi dengan tingkat kematian/aktif tinggi
- **Cluster 2 (Kuning):** Outlier seperti DKI Jakarta dan Jawa Barat

Tabel 3. Dataset *COVID-19* (Cuplikan dari sumber statistik nasional)

IDLocation	Total Cases	Total Recovered	Total Deaths	Total Active Cases	Kematian	Sembuh
DKI Jakarta	208594013	197475361	3420849	7697803	1,639956	94,669717
Jawa Barat	150615019	136201392	2653330	11760297	1,761664	90,430153
Jawa Tengah	104601886	88998394	5687699	9915793	5,437473	85,082973
Jawa Timur	87819795	77904298	6482760	3432737	7,381889	88,709269
Kalimantan Timur	34358373	31284556	1029837	2043980	2,997339	91,05366
Sulawesi Selatan	28476774	26343555	527722	1605497	1,853166	92,508916
Daerah Istimewa Yogyakarta	28145118	24551712	853993	2739413	3,034249	87,232578
Riau	27139120	24923162	772172	1443786	2,845236	91,83482
Banten	26876689	24682144	604657	1589888	2,249745	91,834764
Bali	23785447	21432901	743026	1609520	3,123868	90,109305
Sumatera Barat	20656128	18805743	469542	1380843	2,273137	91,041956
Sumatera Utara	20364282	17594021	605388	2164873	2,972793	86,396471
Kalimantan Selatan	16279357	14621137	540913	1117307	3,322693	89,813971
Sumatera Selatan	12979027	11364511	646452	968064	4,980743	87,560578
Nusa Tenggara Timur	10493075	9359584	224849	908642	2,142832	89,197723
Kalimantan Tengah	10237310	8895952	279728	1061630	2,732437	86,897359
Kepulauan Riau	10187837	9190746	291105	705986	2,857378	90,212927
Lampung	9600051	8132246	631393	836412	6,576975	84,710446
Papua	9418457	6217275	134431	3066751	1,427314	66,011609
Kepulauan Bangka Belitung	8938863	8172432	207668	558763	2,323204	91,425856
Sulawesi Utara	8252633	7103924	264951	883740	3,210503	86,080915
Sulawesi Tengah	7971193	7027539	264951	883740	3,210503	88,161697
Aceh	7895721	6522985	360497	1012239	4,565726	82,614178
Kalimantan Barat	6858126	6308741	149422	399963	2,178758	91,989284
Kalimantan Utara	6598750	5708443	125859	764448	1,907316	86,507945

Nusa Tenggara Barat	6153443	5361080	210638	581725	3,423092	87,123258
Jambi	5456325	4747259	123556	585510	2,264455	87,004696
Sulawesi Tenggara	5333286	4766242	118203	448841	2,216326	89,367831
Papua Barat	5225734	4817872	82576	325286	1,58018	92,195125
Bengkulu	4475334	4034033	97985	343316	2,189446	90,139261
Maluku	3982760	3499078	64634	419048	1,622844	87,855608
Gorontalo	2817730	2569824	92510	155396	3,283139	91,201925
Maluku Utara	2775261	2420767	74370	280124	2,679748	87,226643
Sulawesi Barat	2689823	2395009	63429	231385	2,358111	89,039651



Gambar 3. Hasil Klasterisasi Provinsi Berdasarkan Fitur *COVID-19* (3D Scater Plot)

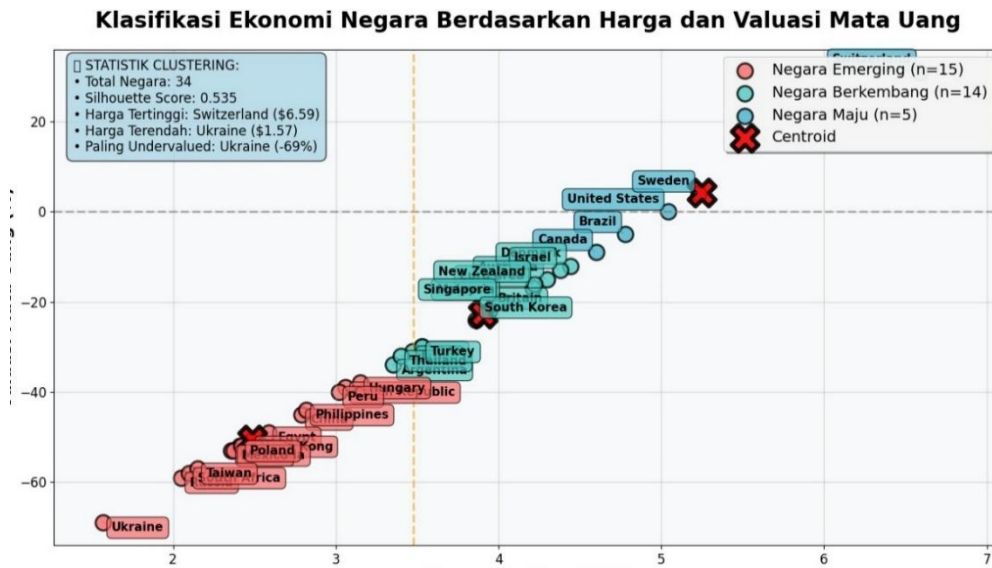
### 3.4. Ilustrasi Konsep *Purchasing Power Parity (Big Mac Index)*

*Big Mac Index* digunakan untuk memetakan nilai tukar berdasarkan daya beli masyarakat di berbagai negara. Hasil klasterisasi negara:

- **Cluster 0 (Merah muda):** Negara emerging, harga *Big Mac* rendah (misal: Ukraina, Indonesia)
- **Cluster 1 (Biru muda):** Negara berkembang, harga menengah (misal: Turki, Thailand)
- **Cluster 2 (Biru tua):** Negara maju, harga tinggi (misal: Swiss, AS)

Tabel 4. Dataset *Bic Mac* dari buku *International Financial Management*

Country	Big Mac Prices		Implied PPP of the dollar	Actual dollar Exchange Rate 07/20/16	Under (-)/over (+) valuation against the dollar, %
	In Local Currency	In Dollars			
United States	\$5.04	5,04	1,00	1,00	0
Argentina	Peso 50	3,35	9,92	14,94	-34
Australia	A\$ 6	4,30	1,17	1,34	-15
Brazil	Real 16	4,78	1,19	3,24	-5
Britain	Pound 3	3,94	0,60	0,76	-22
Canada	C\$ 6	4,60	1,19	1,30	-9
Chile	Peso 2,300	3,53	456,35	651,12	-30
China	Yuan 19	2,79	3,77	6,68	-45
Czech Republic	Koruna 75	3,06	14,88	24,55	-39
Denmark	DK 30	4,44	5,95	6,76	-12
Egypt	Pound 23	2,59	4,56	8,88	-49
Euro area	Euro 4	4,21	0,79	0,91	-17
Hong Kong	HK\$ 19	2,48	3,77	7,76	-51
Hungary	Forint 900	3,15	178,57	285,64	-38
India	Rupee 162	2,41	32,10	67,2	-52
Indonesia	Rupiah 31.000	2,36	6,150.79	13,11250	-53
Israel	Shekel 17	4,38	3,40	3,86	-13
Japan	Yen 370	3,47	73,41	106,73	-31
Malaysia	Ringgit 8	1,99	1,59	4,03	-61
Mexico	Peso 44	2,37	8,73	18,54	-53
New Zealand	NZ\$ 6	4,22	1,19	1,42	-16
Peru	New Sol 10	3,02	1,98	3,31	-40
Philippines	Peso 133	2,82	26,39	47,12	-44
Poland	Zloty 10	2,42	1,98	3,97	-52
Russia	Ruble 130	2,05	25,79	63,41	-60
Singapore	S\$ 5	4,01	0,95	1,36	-20
South Africa	Rand 30	2,10	5,95	14,27	-58
South Korea	Won 4,400	3,86	873,02	1,14095	-24
Sweden	SKr 45	5,23	8,93	8,59	4
Switzerland	SFr 7	6,59	1,39	0,99	31
Taiwan	NT\$ 69	2,15	13,69	32,03	-57
Thailand	Baht 119	3,40	23,61	34,97	-32
Turkey	Lire 11	3,53	2,18	3,04	-30
Ukraine	Hryvnia 39	1,57	7,70	24,80	-69



Gambar 4. Visualisasi Klasterisasi Negara Berdasarkan *Big Mac Index* dan Kurs

### 3.5. Klasterisasi Komponen Neraca Pembayaran Indonesia

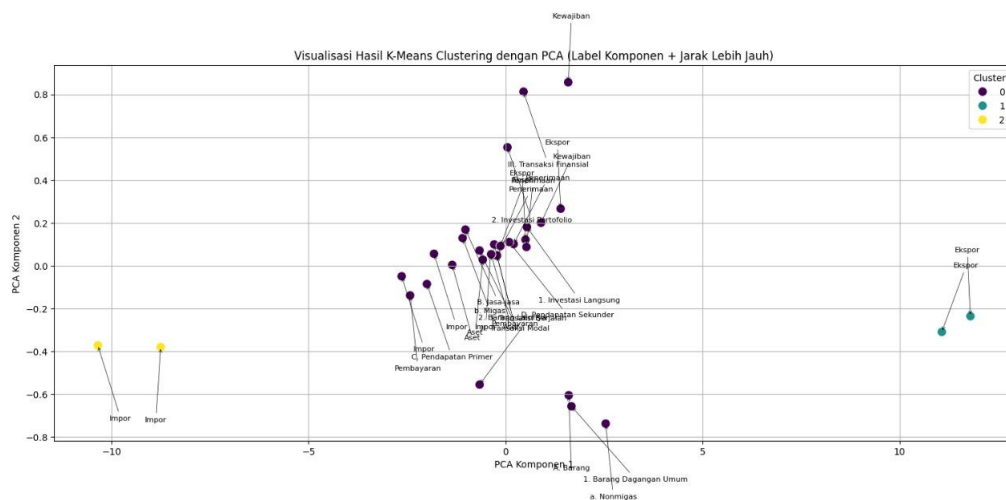
Dataset ini mencakup data neraca pembayaran 2018–2025, termasuk ekspor, impor, transaksi primer/sekunder, dan finansial. Klasterisasi membentuk:

- **Cluster 0 (Ungu):** Komponen defisit
- **Cluster 1 (Kuning):** Komponen fluktuatif dengan nilai besar
- **Cluster 2 (Hijau):** Komponen stabil dan surplus

Tabel 5. Dataset Neraca Pembayaran dari Buku Statistik Keuangan Indonesia

Keterangan	2014	2015	2016	2017	2018
I. Transaksi Berjalan	-27509,87	-17518,74	-16952	-16196	-31560
A. Barang	6982,57	14048,57	15318	18814	431
- Ekspor, fob	175292,79	149124,48	144470	168883	180747
- Impor, fob	-168310,22	-135075,91	-129152	-150069	-175906
B. Jasa	-10009,69	-8696,67	-7084	-7339	-7101
- Ekspor	23530,93	22220,88	23324	25328	27932
- Impor	-33540,62	-30917,55	-30407	-32707	-35034
C. Pendapatan Primer	-29702,61	-28379,13	-29647	-32131	-30420
- Penerimaan	2129,61	2821,7	4048	5575	9182
- Pembayaran	-31832,22	-31200,83	-33695	-37707	-39582
D. Pendapatan Sekunder	5219,87	5508,49	4460	4500	6892
- Penerimaan	9373,54	10361,58	9832	9967	12217
- Pembayaran	-4153,67	-4853,09	-5371	-5467	-5325
II. Transaksi Modal	26,57	16,63	41	46	93
- Penerimaan	26,57	16,63	41	46	93
- Pembayaran	0	0	0	0	0

III. Transaksi Finansial	44916,08	16834,48	29306	28686	25108
1. Investasi Langsung	14733,2	10704,48	16136	18502	13841
2. Investasi Portofolio	26066,63	16182,68	18996	21059	9342
3. Derivatif Finansial	-155,55	19,96	-9	-128	-74
4. Investasi Lainnya	4271,8	-10064,02	-5817	-10747	-1999
IV. Total (I + II + III)	17432,79	-659,01	12394	12536	-5859
V. Selisih Perhitungan Bersih	-2184,2	-439,04	-305	950	-1272
VI. Neraca Keseluruhan (IV+V)	15248,59	-1098,05	12089	11586	-7131
VII. Cad. Devisa dan yg terkait	-15248,59	1098,05	-12089	-11586	-7131
Memorandum:					
- Posisi Cadangan Devisa	111861,6	105931,02	116362	130196	120654
Bln Impor/Bayar Utang LN P	5,45	7,39	8,4	8,3	6,5
- Transaksi Berjalan (% PDB)	-3,09	-2,03	-1,81	-1,6	-2,98
- Rasio Pembayaran Utang (%)	-	-	-	-	-



Gambar 5. Visualisasi PCA dan Klasterisasi Neraca Pembayaran Indonesia

### 3.6. Diskusi

Dari kelima studi di atas, ditemukan bahwa algoritma K-Means mampu memisahkan entitas ekstrem ke dalam kluster tersendiri (contoh: Alibaba, Bus, DKI Jakarta, Ukraina, Ekspor). Ini menunjukkan sensitivitas algoritma terhadap jarak numerik dalam ruang vektor. Kemiripan struktur tiga kluster (rendah-sedang-tinggi) pada semua domain menandakan adanya bentuk alami dalam data numerik multivariabel, yang dapat dimanfaatkan dalam segmentasi strategi, kebijakan publik, maupun pengambilan keputusan bisnis.

Temuan ini menunjukkan bahwa K-Means Clustering tidak hanya bekerja secara efektif pada data dalam satu domain, tetapi juga mampu menunjukkan pola klasifikasi alami pada konteks yang sangat berbeda. Konsistensi dalam pemisahan kluster berdasarkan jarak numerik memperkuat posisi K-Means sebagai algoritma eksploratif yang handal untuk segmentasi data multivariabel.

## 4. Kesimpulan

Berdasarkan hasil klasterisasi lima dataset dari sektor ekonomi dan kesehatan, dapat disimpulkan bahwa algoritma K-Means Clustering berhasil mengelompokkan entitas ke dalam struktur yang bermakna secara numerik. Klasterisasi pada data IPO mampu membedakan perusahaan dengan valuasi ekstrem (Alibaba) dari perusahaan lain. Pada data kendaraan, segmentasi berdasarkan konsumsi oksigen memberikan gambaran jelas tentang dampak

lingkungan masing-masing jenis kendaraan. Untuk data COVID-19, klusterisasi membantu mengidentifikasi wilayah yang menjadi pusat kasus tinggi, serta wilayah dengan karakteristik menengah dan rendah. Analisis Big Mac Index membuktikan konsep *Purchasing Power Parity* dapat direpresentasikan secara efektif melalui pemetaan harga barang global. Sementara itu, klusterisasi komponen neraca pembayaran menunjukkan pemisahan antara komponen defisit, fluktuatif, dan stabil. Secara umum, pola tiga kluster yang konsisten menunjukkan bahwa K-Means sensitif terhadap distribusi data dalam ruang multidimensi dan dapat menjadi alat eksploratif yang kuat untuk mendukung pengambilan keputusan berbasis data. Penelitian ini menunjukkan bahwa penerapan algoritma kluster dapat diadopsi lintas domain untuk menyederhanakan kompleksitas data dan menemukan struktur alami dalam berbagai konteks.

## Referensi

1. J. Han, J. Pei, and M. Kamber, *Data Mining: Concepts and Techniques*, 4th ed. Cambridge, MA: Morgan Kaufmann, 2022.
2. C. S. Eun and B. G. Resnick, *International Financial Management*, 9th ed. New York: McGraw-Hill Education, 2021.
3. R. H. Friis, *Essentials of Environmental Health*, 4th ed. Burlington, MA: Jones & Bartlett Learning, 2023.
4. Badan Pusat Statistik, *Neraca Pembayaran Indonesia 2018–2023*, Jakarta: BPS RI, 2023. [Online]. Available: <https://www.bps.go.id/publication>
5. R. A. Jalil, M. M. A. Rahman, and M. K. Hasan, "Forecasting and clustering of COVID-19 confirmed cases for different regions using hybrid machine learning approach," *Sustainable Cities and Society*, vol. 78, p. 103541, Apr. 2022. doi: <https://doi.org/10.1016/j.scs.2021.103541>