



## ANALISIS KOMPARATIF K-MEANS UNTUK SEGMENTASI DATA EKONOMI, SOSIAL, DAN KEUANGAN

Dahlia Agustina<sup>1</sup>, Dennis Putra Nanda<sup>2</sup>, Fahri Fadhlurrahman Rendy Dhaif Muharram<sup>3</sup>, Moh Adhyaksa Pratama<sup>4</sup>, Yunita Rahma Anggraeni<sup>5</sup>, Zurnan Alfian<sup>6</sup>  
<sup>123456</sup>Fakultas Ilmu Komputer, Universitas Pamulang  
 email: [dahliaagustina2323@gmail.com](mailto:dahliaagustina2323@gmail.com)<sup>1</sup>, [dennis.pn23@gmail.com](mailto:dennis.pn23@gmail.com)<sup>2</sup>, [dhaifmuharram@gmail.com](mailto:dhaifmuharram@gmail.com)<sup>3</sup>, [adhyaksaaryansyah@gmail.com](mailto:adhyaksaaryansyah@gmail.com)<sup>4</sup>, [yunrhm10@mail.com](mailto:yunrhm10@mail.com)<sup>5</sup>

Informasi Artikel	ABSTRACT
<p>Riwayat artikel :                      Disubmit : 28 Juni 2025                      Direvisi : 1 Desember 2025                      Diterima : 28 Desember 2025                      Dipublikasi : 31 Desember 2025</p> <p><b>Keywords:</b>                      K-Means clustering, data segmentation, comparative analysis, data standardization, multidisciplinary study.</p>	<p><i>This study aims to analyse the comparative application of the K-Means algorithm in five case studies from the economic, social, and financial domains. The case studies used include poverty data in Indonesia, financial reports of state-owned enterprises in the transportation sector, tobacco consumption in the United States, cross-country education budget allocation, and foreign exchange rates against the Rupiah. The K-Means algorithm was applied with appropriate pre-processing steps, such as data standardisation and determining the optimal number of clusters using the Elbow Method in certain cases. The datasets were obtained from academic literature, official financial reports, and international institution publications. The analysis was conducted using a quantitative approach to produce clusters, which were then interpreted descriptively and qualitatively. The results of the study show that K-Means is able to group data effectively in the five cases analysed, especially when followed by the appropriate methodological steps. The resulting clusters facilitate understanding of data patterns, both for evaluating socio-economic policies and for supporting financial decision-making. The conclusion of this study is that K-Means is a flexible and versatile data segmentation tool, but its successful implementation is highly dependent on the data context, the methodology applied, and careful interpretation of the results.</i></p>
	ABSTRAK
<p><b>Kata Kunci:</b>                      K-Means clustering, segmentasi data, analisis komparatif, standarisasi data, studi multidisiplin.</p>	<p>Penelitian ini bertujuan untuk menganalisis penerapan algoritma K-Means secara komparatif pada lima studi kasus dari domain ekonomi, sosial, dan keuangan. Studi kasus yang digunakan meliputi data kemiskinan di Indonesia, laporan keuangan BUMN sektor perhubungan, konsumsi tembakau di Amerika Serikat, alokasi anggaran pendidikan lintas negara, dan kurs valuta asing terhadap Rupiah. Algoritma K-Means diterapkan dengan langkah pra-pemrosesan yang sesuai, seperti standarisasi data dan penentuan jumlah kluster optimal menggunakan Elbow Method pada kasus tertentu. Dataset diperoleh dari literatur akademis, laporan keuangan resmi, dan publikasi lembaga internasional. Analisis dilakukan dengan pendekatan kuantitatif untuk menghasilkan kluster, yang kemudian diinterpretasikan secara deskriptif kualitatif. Hasil penelitian menunjukkan bahwa K-Means mampu mengelompokkan data secara efektif pada kelima kasus yang dianalisis, terutama ketika diikuti dengan langkah metodologi yang tepat. Kluster yang dihasilkan mempermudah pemahaman pola data, baik untuk evaluasi kebijakan sosial-ekonomi maupun untuk mendukung pengambilan keputusan keuangan. Kesimpulan dari studi ini adalah bahwa K-Means merupakan alat segmentasi data yang fleksibel dan serbaguna, namun keberhasilan implementasinya sangat bergantung pada konteks data, metodologi yang diterapkan, dan interpretasi hasil yang cermat.</p>





## PENDAHULUAN

Perkembangan teknologi informasi dan transformasi digital telah mendorong peningkatan volume data yang bersifat besar, kompleks, dan heterogen pada berbagai sektor, khususnya dalam bidang ekonomi, sosial, dan keuangan. Kondisi ini menuntut pemanfaatan teknik analisis data berbasis komputasi yang mampu mengidentifikasi pola dan struktur data secara objektif guna mendukung pengambilan keputusan berbasis data (*data-driven decision making*). Namun demikian, perbedaan skala variabel, karakteristik data lintas domain, serta kompleksitas struktur data sering kali menjadi tantangan utama dalam proses analisis dan interpretasi data secara efektif (Kurniawan & Defit, 2020; Bachtiar et al., 2023).

Salah satu metode yang banyak digunakan untuk mengatasi permasalahan tersebut adalah algoritma *K-Means Clustering*, yaitu teknik *unsupervised learning* yang bertujuan mengelompokkan data ke dalam sejumlah kluster berdasarkan tingkat kemiripan karakteristiknya menggunakan pusat kluster (*centroid*). Algoritma *K-Means* dikenal memiliki keunggulan berupa kesederhanaan konsep, efisiensi komputasi, serta kemudahan implementasi pada data numerik berskala besar, sehingga banyak diterapkan dalam berbagai penelitian berbasis eksplorasi data (Larasati et al., 2021).

Berbagai penelitian empiris menunjukkan efektivitas algoritma *K-Means* dalam konteks domain tertentu. Sano dan Nindito (2016) memanfaatkan *K-Means* untuk mengelompokkan tingkat kemiskinan antarprovinsi di Indonesia berdasarkan indikator sosial-ekonomi dan berhasil mengidentifikasi pola kemiskinan regional yang relevan bagi perumusan kebijakan. Putri dan Rahmah (2024) menerapkan *K-Means* dalam segmentasi nasabah perusahaan asuransi untuk mendukung pengambilan keputusan strategis, sementara Irawan et al. (2023) menunjukkan bahwa *K-Means* efektif dalam mengelompokkan nasabah perbankan berdasarkan profil transaksi keuangan.

Selain penerapan pada domain spesifik, beberapa penelitian juga menyoroti aspek metodologis dalam penggunaan *K-Means*. Kurniawan dan Defit (2020) menegaskan bahwa proses pra-pemrosesan data, khususnya standardisasi variabel, berpengaruh signifikan terhadap kualitas hasil klusterisasi pada data dengan skala yang berbeda. Larasati et al. (2021) menambahkan bahwa pemilihan jumlah kluster optimal melalui *Elbow Method* dapat meningkatkan validitas dan interpretabilitas kluster yang dihasilkan. Studi lain membandingkan *K-Means* dengan algoritma klusterisasi alternatif seperti *DBSCAN* dan *Hierarchical Clustering*, dan menemukan bahwa meskipun *K-Means* unggul dari sisi efisiensi komputasi, algoritma ini sensitif terhadap karakteristik data dan inisialisasi *centroid* (Bachtiar et al., 2023; Sufairoh et al., 2023).





Meskipun literatur menunjukkan keberhasilan K-Means dalam berbagai konteks, sebagian besar penelitian terdahulu masih terbatas pada satu domain atau satu jenis data tertentu. Penelitian-penelitian tersebut umumnya belum membahas secara komparatif bagaimana algoritma K-Means bekerja pada data lintas disiplin yang memiliki karakteristik berbeda, seperti data ekonomi makro, laporan keuangan, data deret waktu, dan data lintas negara. Akibatnya, pemahaman mengenai fleksibilitas, konsistensi, serta generalisasi algoritma K-Means dalam konteks multidisiplin masih relatif terbatas.

Berdasarkan kondisi tersebut, terdapat kesenjangan penelitian yang berkaitan dengan minimnya kajian komparatif lintas domain serta terbatasnya pembahasan mengenai pengaruh tahapan metodologis, seperti pra-pemrosesan data, standardisasi fitur, dan pemilihan jumlah kluster, terhadap kualitas dan interpretabilitas hasil klusterisasi pada data multidisiplin. Kesenjangan ini menunjukkan perlunya penelitian yang tidak hanya berfokus pada hasil klusterisasi, tetapi juga pada proses metodologis yang mendasarinya.

Oleh karena itu, kebaruan penelitian ini terletak pada analisis komparatif penerapan algoritma K-Means pada lima studi kasus lintas domain yang mencakup data ekonomi, sosial, dan keuangan dengan karakteristik yang beragam. Penelitian ini tidak hanya mengevaluasi pola kluster yang dihasilkan, tetapi juga mengkaji praktik metodologis yang digunakan dalam setiap konteks data, sehingga memberikan kontribusi empiris dan metodologis yang belum banyak dikaji dalam penelitian sebelumnya.

Penelitian ini bertujuan untuk menganalisis penerapan algoritma K-Means secara komparatif pada berbagai jenis data guna mengevaluasi efektivitas, fleksibilitas, dan interpretabilitas hasil klusterisasi. Melalui pendekatan ini, penelitian diharapkan mampu memberikan pemahaman yang lebih komprehensif mengenai penerapan K-Means pada data multidisiplin, sekaligus menjadi referensi metodologis bagi peneliti dan pengambil kebijakan dalam memanfaatkan teknik klusterisasi sebagai alat analisis eksploratif yang mendukung pengambilan keputusan berbasis data.

## METODE PENELITIAN

Penelitian ini menggunakan pendekatan studi kasus komparatif dengan lima dataset berbeda yang diambil dari berbagai sumber literatur. Dataset pertama adalah data kemiskinan di Indonesia yang bersumber dari buku *Perekonomian Indonesia* karya Tulus T.H. Tambunan, memuat data deret waktu mengenai tingkat kemiskinan dari tahun 1976 hingga 2017. Dataset kedua adalah laporan keuangan BUMN sektor perhubungan yang diambil dari buku *Kebangkitan BUMN Sektor Perhubungan* yang memuat data keuangan PT Angkasa Pura II tahun 2010 hingga 2012. Dataset ketiga adalah data konsumsi tembakau per kapita di Amerika Serikat, bersumber dari buku *Modern Business Statistics* karya Freund & Williams. Dataset keempat adalah data alokasi anggaran pendidikan lintas negara, yang



This work is licensed under a [Creative Commons Attribution-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/)



diambil dari *World Development Report 1988* terbitan Bank Dunia. Dataset kelima adalah data kurs valuta asing per 1 November 2016 yang diambil dari situs [seputarforex.com](http://seputarforex.com). Analisis dilakukan dengan menerapkan algoritma *K-Means Clustering* pada masing-masing dataset. Tahapan analisis diawali dengan pra-pemrosesan data, yang meliputi standardisasi fitur numerik menggunakan *StandardScaler* untuk menyamakan skala variabel. Pada dataset kurs valuta asing, dilakukan *feature engineering* dengan menambahkan variabel *spread* untuk menangkap perbedaan antara kurs jual dan kurs beli. Penentuan jumlah kluster ( $k$ ) dilakukan menggunakan Elbow Method dengan menghitung *Within-Cluster Sum of Squares* (WCSS) dan memvisualisasikan grafik elbow. Setelah jumlah kluster ditentukan, algoritma K-Means diimplementasikan dengan menggunakan pustaka *Scikit-learn* pada Python. Hasil klusterisasi divisualisasikan dengan *scatter plot* dan grafik kluster menggunakan *Matplotlib* dan *Seaborn*. Analisis hasil dilakukan secara deskriptif kualitatif untuk menginterpretasi pola kluster pada setiap dataset. Pseudocode algoritma K-Means yang diterapkan dapat dituliskan sebagai berikut:

## HASIL DAN PEMBAHASAN

### 1. Perekonomian Indonesia 1965 2028

Tabel 5.1: Kemiskinan di Indonesia, 1976–2017

Tahun	Jumlah orang miskin (juta)			Persentase kemiskinan (%)		
	Kota	Desa	Jumlah	Kota	Desa	Jumlah
1976	10,00	44,20	54,20	38,79	40,37	40,08
1980	9,50	32,80	42,30	29,04	28,42	28,56
1984	9,30	25,70	35,00	23,14	21,18	21,64
1987	9,70	20,30	30,00	20,14	16,14	17,42
1990	9,40	17,80	27,20	16,75	14,33	15,08
1996	9,42	24,59	34,01	13,38	19,78	17,47
1998	17,60	31,90	49,50	21,92	25,72	24,23
1999	15,64	32,33	47,97	19,41	26,03	23,43
2000	12,30	28,44	38,74	14,60	22,38	19,14
2001	8,60	29,27	37,87	9,76	24,84	18,41
2002	13,30	25,09	38,39	14,46	21,10	18,20
2003	12,20	25,14	37,34	13,57	20,23	17,42
2004	11,30	24,85	36,15	12,13	20,11	16,66

Angka di Tabel 5.1 didasarkan pada metode indeks head count (HC), yang umum digunakan untuk menghitung penduduk miskin. Menurut metode ini, penduduk miskin adalah penduduk yang berada di bawah suatu batas, yang disebut garis kemiskinan.

Gambar 1. Kemiskinan di Indonesia, 1976-2017

Dataset ini berisi data historis mengenai tingkat kemiskinan di Indonesia dari tahun 1976 hingga 2017. Data bersumber dari buku "Perekonomian Indonesia" karya Tulus T.H. Tambunan, yang mengolah data dari Badan Pusat Statistik (BPS). Data mencakup jumlah absolut orang miskin (dalam juta jiwa) dan persentase kemiskinan. Data ini dibedakan menjadi tiga kategori utama: kemiskinan di wilayah Perkotaan (Kota), Perdesaan (Desa), serta gabungan keduanya (Jumlah). Seluruh nilai dalam satuan juta jiwa atau persen (%).

#### a. Hasil Klusterisasi

- **Cluster 0:** Tahun-tahun dengan tingkat kemiskinan **sangat tinggi**, terutama pada periode awal (misalnya 1976, 1980) dan saat terjadi krisis ekonomi (1998-1999).



This work is licensed under a [Creative Commons Attribution-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/).

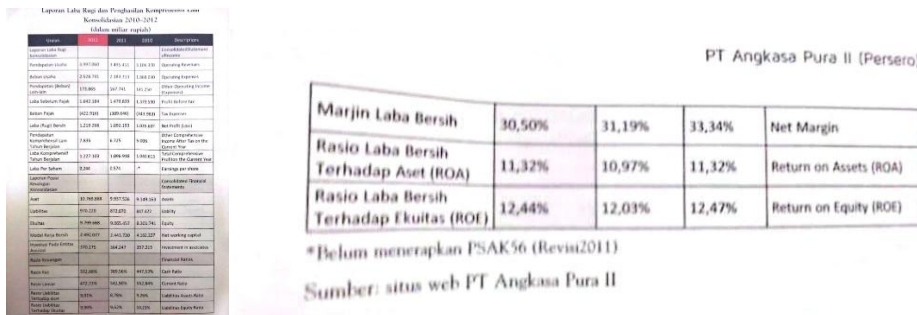


- **Cluster 1:** Tahun-tahun dengan tingkat kemiskinan **menengah**, yang mencerminkan periode transisi atau pasca-perbaikan ekonomi (misalnya awal 1990-an dan awal 2000-an).
- **Cluster 2:** Tahun-tahun dengan tingkat kemiskinan **relatif rendah**, terutama pada periode lebih modern (misalnya setelah 2010), di mana angka kemiskinan sudah lebih terkendali.

**b. Kesimpulan**

Nilai K-Means Clustering berhasil mengelompokkan data tahunan ke dalam periode-periode yang memiliki profil kemiskinan yang serupa. Hasil klasterisasi ini membantu mengidentifikasi era-era penting dalam sejarah penanggulangan kemiskinan di Indonesia serta menunjukkan tren perubahan struktural dari waktu ke waktu. kategori kecil yang dapat diberi perhatian khusus untuk efisiensi anggaran ke depan.

**2. Kebangkitan BUMN Perhubungan**



Gambar 2. Laporan Laba Rugi dan Penghasilan Komprehensif Lain Konsolidasian 2010-2012

Dataset data Laporan Keuangan Konsolidasian PT Angkasa Pura II untuk periode tahun 2010 hingga 2012, yang bersumber dari buku "Kebangkitan BUMN Sektor Perhubungan". Data yang dianalisis mencakup berbagai metrik dari laporan laba rugi seperti Pendapatan Usaha dan Laba Bersih, laporan posisi keuangan seperti Aset dan Liabilitas, serta berbagai rasio keuangan penting termasuk ROA, ROE, dan Net Margin. Seluruh nilai moneter disajikan dalam satuan miliar rupiah.

**a. Hasil Klasterisasi**

- Cluster 0: Sektor dominan seperti Industri Pengolahan dan Perdagangan.
- Cluster 1: Sektor menengah dan stabil seperti Bangunan, Jasa-jasa, dan Transportasi.
- Cluster 2: Sektor kecil seperti Listrik, Kehutanan, dan Perikanan.

**b. Kesimpulan**

K-Means berhasil mengelompokkan sektor berdasarkan skala kontribusinya. Hasil ini membantu identifikasi sektor dominan dan sektor yang butuh perhatian lebih.



This work is licensed under a [Creative Commons Attribution-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/)



### 3. Statistika Untuk Ekonomi Dan Niaga 1 Edisi Baru

DAFTAR IV (9)  
PEMAKAIAN TEMBAKAU PER-KAPITA  
DI AMERIKA SERIKAT SELAMA 1930-1965  
(dalam pounds)

Tahun	Pemakaian	Tahun	Pemakaian
1930	8,8	1943	11,5
1931	8,4	1944	11,2
1932	7,8	1945	12,5
1933	7,8	1946	12,2
1934	8,3	1947	12,0
1935	8,2	1948	12,1
1936	8,8	1949	11,9
1937	9,0	1950	12,0
1938	8,8	1951	12,5
1939	8,8	1952	12,9
1940	9,1	1953	12,9
1941	9,0	1954	12,2
1942	10,7	1955	12,3

Sumber : Modern Business Statistics, oleh: Freund & Williams.

Gambar 3. Pemakaian Tembakau Per-Kapita Di Amerika Serikat Selama 1930-1965

Dataset ini menyajikan data historis mengenai pemakaian tembakau per-kapita di Amerika Serikat untuk periode tahun 1930 hingga 1955. Data ini bersumber dari buku "Modern Business Statistics" oleh Freund & Williams, yang juga direplikasi dalam literatur statistika lainnya. Dataset ini bersifat univariat, hanya terdiri dari satu fitur utama yaitu nilai "Pemakaian" tahunan yang diukur dalam satuan pounds.

#### a. Hasil Klasterisasi

- Klasterisasi dengan  $k=2$  secara efektif membagi periode waktu menjadi dua kelompok yang berbeda.
- Klaster 0: Berisi tahun-tahun dengan tingkat pemakaian rendah, umumnya mencakup periode awal dari tahun 1930-an hingga awal 1940-an.
- Klaster 1: Berisi tahun-tahun dengan tingkat pemakaian tinggi, yang dimulai sekitar tahun 1942 dan berlanjut hingga akhir periode data di tahun 1955. Tingkat signifikan dari 36% (2000) menjadi 53% (2005), mencerminkan penguatan struktur modal internal.

#### b. Kesimpulan

K-Means berhasil melakukan periodisasi historis dengan memisahkan era pemakaian rendah dan era pemakaian tinggi secara jelas. Meskipun dataset sangat sederhana, metode ini terbukti efektif dalam menemukan titik perubahan (turning point) atau pergeseran tren dalam data deret waktu.





#### 4. Analisis Data Alokasi Pendidikan Tahun 1972 & 1986

MEMBAH KEMERDEKAAN GLOBAL

**Tabel 6.1 Alokasi untuk pendidikan dalam % dari anggaran pemerintah pusat**

Negara	1972	1986	Negara	1972	1986
Argentina	20.0	6.0	Malawi	15.6	11.0
Burkina Faso	20.6	17.7	Mekiko	16.4	11.5
Bangladesh	14.8	9.9	Maroko	19.2	16.6
Bolivia	31.3	11.6	Sierra Leone	15.0	10.9
Brazil	8.3	3.0	Sri Lanka	13.0	6.4
Burma	13.0	11.7	Swedia	14.8	8.9
Chili	14.3	12.5	Tanzania	17.3	7.2
Kosta Rika	28.3	16.2	Tunisia	20.5	14.3
Denmark	16.0	9.2	Turki	18.1	11.9
India	16.7	7.2	Zambia	19.0	16.0
Lesotho	22.4	15.5	Amerika Serikat	3.2	1.7

Sumber: Bank Dunia, World Development Report 1988, Oxford University Press, Inc. 266-7.

Gambar 4. Alokasi Untuk Pendidikan Dalam % Dari Anggaran Pemerintah Pusat

Dataset ini menyajikan data komparatif mengenai alokasi anggaran untuk sektor pendidikan sebagai persentase dari total anggaran pemerintah pusat. Data ini bersumber dari laporan Bank Dunia (World Development Report 1988) dan mencakup berbagai negara di dunia. Analisis berfokus pada dua fitur utama, yaitu persentase alokasi anggaran pada tahun 1972 dan tahun 1986, untuk melihat perubahan dan pola kebijakan antar negara.

##### a. Hasil Klasterisasi

- Klasterisasi dengan  $k=3$  berhasil mengelompokkan negara ke dalam tiga kategori yang berbeda secara jelas.
- Klaster 0: Negara-negara dengan alokasi anggaran pendidikan yang cenderung rendah di kedua tahun.
- Klaster 1: Negara-negara dengan alokasi anggaran yang sangat tinggi di kedua tahun (misalnya Bolivia, Kosta Rika).
- Klaster 2: Negara-negara dengan alokasi anggaran menengah yang cenderung stabil atau mengalami penurunan moderat.

##### b. Kesimpulan

K-Means Dengan demikian, dapat disimpulkan bahwa K-Means Clustering secara efektif berhasil mengidentifikasi dan mengelompokkan negara-negara yang memiliki kebijakan fiskal pendidikan serupa. Analisis ini tidak hanya menunjukkan pola pengelompokan, tetapi juga memberikan wawasan berharga mengenai perbedaan prioritas pembangunan antar negara serta bagaimana prioritas tersebut dapat berubah dari waktu ke waktu.





## 5. Kurs Valuta Asing Per. 1 November 2026

Tabel 10.1 Kurs Valuta Asing terhadap Rupiah  
Per 1 November 2016

Simbol	Mata Uang	Kurs Jual per unit	Kurs Beli Per unit
AUD	Australia - Dollar	10.146,90	9.863,90
CAD	Canada - Dollar	9.861,70	9.599,70
CHF	Swiss - Franc	13.398,15	13.054,15
CNY	China - Yuan	2.006,00	1.840,30
DKK	Denmark - Krona	1.963,90	1.899,50
EUR	Europa - Euro	14.526,60	14.156,60
GBP	English - Poundsterling	16.194,80	15.749,80
HKD	HongKong - Dollar	1.697,75	1.663,75
JPY	Japan - Yen	128,77	121,83
JPY	Japan - Yen	9.501,90	9.207,90
NZD	New Zealand - Dollar	3.527,45	3.423,45
SAR	Riyal - Saudi	1.482,05	1.427,75
SEK	Swedia - Krona	9.709,72	9.341,72
SGD	Singapura - Dollar	13,00	12.884,00
USD	Amerika - USD Dollar		

Sumber: <http://www.seputarforex.com/da>

Gambar 5. Kurs Valuta Asing Terhadap Rupiah Per 1 November 2016

Dataset ini menyajikan data kurs berbagai valuta asing terhadap Rupiah. Data ini bersifat *cross-sectional*, yang mencerminkan kondisi pasar pada satu titik waktu spesifik, yaitu 1 November 2016. Bersumber dari situs *seputarforex.com*, dataset ini mencakup berbagai mata uang dunia dan memiliki dua fitur utama: 'Kurs Jual per unit' dan 'Kurs Beli per unit', yang keduanya dinilai dalam Rupiah.

### a. Tujuan

Mengelompokkan mata uang asing ke dalam kategori-kategori berdasarkan kemiripan karakteristik nilai tukarnya (level harga dan spread) menggunakan K-Means Clustering, dan membuat fitur baru yaitu 'Spread' (selisih antara Kurs Jual dan Kurs Beli) untuk menangkap volatilitas atau biaya transaksi.

- Klaster 1 (Mata Uang Bernilai Rendah): Berisi mata uang dengan nilai tukar rendah, seperti Yen Jepang (JPY) dan Yuan China (CNY).
- Klaster 2 (Mata Uang Bernilai Menengah): Berisi mata uang utama lainnya dengan nilai tukar di tingkat menengah, seperti Dolar Amerika (USD) dan Dolar Australia (AUD).

### b. Hasil dan Interpretasi

Berdasarkan hasil analisis, K-Means Clustering terbukti efektif dalam segmentasi pasar valuta asing, mengelompokkan mata uang ke dalam kategori nilai yang jelas (tinggi, menengah, dan rendah). Pengelompokan ini valid secara statistik dan intuitif secara ekonomi, memberikan wawasan cepat bagi investor atau analis keuangan untuk memahami struktur pasar mata uang terhadap Rupiah.

### c. Kesimpulan

Hasil analisis menunjukkan bahwa K-Means Clustering efektif dalam segmentasi pasar valuta asing, mengelompokkan mata uang ke dalam kategori nilai yang jelas (tinggi, menengah, dan rendah). Pengelompokan ini valid secara statistik dan intuitif secara ekonomi, memberikan wawasan cepat bagi investor atau analis keuangan untuk memahami struktur dan stratifikasi pasar mata uang terhadap Rupiah.



This work is licensed under a [Creative Commons Attribution-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/)



#### d. Dataset Kemiskinan di Indonesia

No	Tahun	Kemiskinan Kota (%)	Kemiskinan Desa (%)
1	1976	10.0	44.2
2	1980	9.5	32.8
3	1984	9.3	25.7
4	1987	9.7	20.3
5	1990	9.4	17.8
6	1996	9.42	24.59
7	1998	17.6	31.9
8	1999	15.84	32.33
9	2000	12.3	26.44
10	2001	8.6	29.27
11	2002	13.3	25.09
12	2003	12.2	25.14
13	2004	11.3	24.85

Dataset ini merupakan data deret waktu (*time-series*) dari tahun 1976 hingga 2004, yang bersumber dari buku *Perekonomian Indonesia* karya Tulus T.H. Tambunan, dengan data Survei Sosial Ekonomi Nasional (Susenas) dari Badan Pusat Statistik (BPS). Fitur utama yang dianalisis adalah jumlah penduduk miskin (dalam juta jiwa) yang dibedakan berdasarkan wilayah perkotaan (Kota) dan perdesaan (Desa). Tujuan analisis ini adalah untuk mengidentifikasi dan mengelompokkan periode-periode waktu dengan profil kemiskinan serupa.

#### e. Dataset Keuangan BUMN

No	Tahun	Pendapatan Usaha	Laba Bersih
1	2010	3.106.370	1.035.607
2	2011	3.495.411	1.090.193
3	2012	3.997.060	1.219.268

Dataset ini berisi laporan keuangan tahunan PT Angkasa Pura II untuk periode 2010 hingga 2012, yang diekstraksi dari buku *Kebangkitan BUMN Sektor Perhubungan* oleh Cyrillus Harinowo. Variabel yang digunakan mencakup metrik keuangan penting seperti Pendapatan Usaha, Laba Bersih, Aset, Ekuitas, dan rasio keuangan lainnya. Dataset ini bertujuan untuk mengelompokkan kinerja perusahaan dari tahun ke tahun.

#### f. Dataset Pemakaian Tembakau

No	Tahun	Pemakaian Tembakau
1	1930	8.8
2	1931	8.4
3	1932	7.6
4	1933	7.8
5	1934	8.3
6	1935	8.2
7	1936	8.8
8	1937	9.0
9	1938	8.8





10	1939	8.8
11	1940	9.1
12	1941	9.0
13	1942	10.7
14	1943	11.5
15	1944	11.2
16	1945	12.5
17	1946	12.2
18	1947	12.0
19	1948	12.1
20	1949	11.9
21	1950	12.0
22	1951	12.5
23	1952	12.9
24	1953	12.9
25	1954	12.2
26	1955	12.3

Dataset ini adalah data deret waktu historis yang menunjukkan tren pemakaian tembakau per kapita di Amerika Serikat antara 1930–1955, bersumber dari *Modern Business Statistics* oleh Freund & Williams. Dengan satu variabel utama, 'Pemakaian' dalam satuan pounds, dataset ini menjadi studi kasus menarik untuk menguji kemampuan K-Means dalam melakukan periodisasi berdasarkan satu fitur.

#### g. Dataset Alokasi Anggaran Pendidikan

No	Entitas	Alokasi 1972 (%)	Alokasi 1986 (%)
1	Argentina	20.0	6.0
2	Burkina Faso	20.6	17.7
3	Bangladesh	14.8	9.9
4	Bolivia	31.3	11.6
5	Brasil	8.3	3.0
6	Burma	15.0	11.7
7	Chili	14.3	12.5
8	Kosta Rika	28.3	16.2
9	Denmark	16.0	9.2
10	Italia	16.7	7.2
11	Lesotho	22.4	15.5
12	Malawi	15.8	11.0
13	Meksiko	16.4	11.5
14	Maroko	19.2	16.6
15	Selandia Baru	16.9	10.9
16	Sri Lanka	13.0	8.4
17	Swedia	14.8	8.9
18	Tanzania	17.3	7.2
19	Tunisia	30.5	14.3
20	Turki	18.1	11.9
21	Zambia	19.0	16.0
22	Amerika Serikat	3.2	1.7





Dataset ini bersifat lintas-seksi (cross-sectional), membandingkan alokasi anggaran sektor pendidikan di berbagai negara pada 1972 dan 1986, dengan sumber data dari laporan Bank Dunia (*World Development Report 1988*). Fitur yang dianalisis adalah persentase alokasi dana pendidikan dari total anggaran pemerintah pusat, bertujuan untuk mengelompokkan negara-negara berdasarkan kesamaan prioritas kebijakan fiskal di bidang pendidikan.

#### h. Dataset Kurs Valuta Asing

No	Mata Uang	Kurs Jual	Kurs Beli
1	AUD	10.146,90	9.863,90
2	CAD	9.861,70	9.599,70
3	CHF	13.398,15	13.054,15
4	CNY	2.006,00	1.840,30
5	DKK	1.963,90	1.899,50
6	EUR	14.526,60	14.156,60
7	GBP	16.194,80	15.749,80
8	HKD	1.697,75	1.663,75
9	JPY	126,77	121,83
10	NZD	3.527,45	3.423,45
11	SAR	1.482,05	1.427,75
12	SEK	9.420,72	9.341,72
13	SGD	13.184,00	12.884,00
14	USD	13.075,00	13.055,00

Dataset ini juga bersifat cross-sectional yang merekam kondisi pasar pada satu waktu spesifik, yaitu 1 November 2016. Data ini *bersumber* dari situs seputarforex.com dan dikutip dalam buku "Manajemen Keuangan Fundamental". Fitur utamanya adalah Kurs Jual dan Kurs Beli dari berbagai mata uang dunia terhadap Rupiah (IDR). Analisis pada dataset ini bertujuan untuk mengelompokkan mata uang ke dalam kategori-kategori berdasarkan nilai tukarnya.

#### 6. Implementasi K-Means Berbasis Python: Alur Kerja dan Visualisasi

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.cluster import KMeans
from sklearn.preprocessing import StandardScaler
import os

sns.set_theme(style="whitegrid")
NAMA_FOLDER_OUTPUT = 'hasil_visualisasi_ringkas'

if not os.path.exists(NAMA_FOLDER_OUTPUT):
    os.makedirs(NAMA_FOLDER_OUTPUT)
    print(f"Folder '{NAMA_FOLDER_OUTPUT}' telah dibuat.")

def analisis_dan_plot_klaster(config):
    """
    Fungsi generik untuk menjalankan analisis kluster KMeans dan membuat visualisasi.
    """
```





```

print(f"--- Menganalisis: {config['judul']} ---")

df = pd.DataFrame(config['data'])

if config.get('pra_pemrosesan'):
    df = config['pra_pemrosesan'](df)

fitur_untuk_klaster = df[config['kolom_fitur']]

if config.get('gunakan_skala', False):
    fitur_scaled = StandardScaler().fit_transform(fitur_untuk_klaster)
else:
    fitur_scaled = fitur_untuk_klaster

kmeans = KMeans(n_clusters=config['n_clusters'], random_state=42, n_init=10)
df['Klaster'] = kmeans.fit_predict(fitur_scaled)

plt.figure(figsize=config['ukuran_plot'])
ax = sns.scatterplot(
    data=df,
    x=config['kolom_x'],
    y=config['kolom_y'],
    hue='Klaster',
    palette=config['palette'],
    s=150,
    edgecolor='black'
)

for i, txt in enumerate(df[config['kolom_label']]):
    ax.annotate(txt, (df[config['kolom_x']].iloc[i], df[config['kolom_y']].iloc[i]),
                textcoords="offset points", xytext=(0, 8), ha='center', fontsize=9)

plt.title(config['judul_plot'], fontsize=16)
nama_file = os.path.join(NAMA_FOLDER_OUTPUT, config['nama_file'])
plt.savefig(nama_file, bbox_inches='tight')
plt.close()
print(f"Grafik telah disimpan sebagai: {nama_file}\n")

```

#### a. Lingkungan Pengembangan dan Pustaka *Python*

Analisis data dalam penelitian ini dilakukan menggunakan bahasa pemrograman Python (versi 3.x) dalam skrip mandiri untuk memastikan objektivitas dan efisiensi. Pustaka utama yang digunakan dalam penelitian ini meliputi:

- **Pandas:** Digunakan untuk strukturisasi dan manipulasi data, dengan setiap dataset diubah menjadi objek DataFrame untuk memudahkan pemrosesan.
- **Scikit-learn (sklearn):** Pustaka utama untuk implementasi machine learning, menggunakan modul `sklearn.cluster.KMeans` untuk K-Means Clustering dan `sklearn.preprocessing.StandardScaler` untuk pra-pemrosesan data.





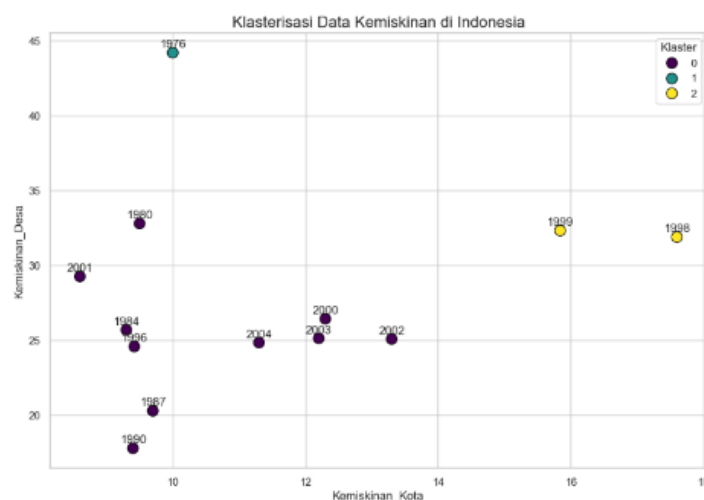
- **Matplotlib & Seaborn:** Kedua pustaka ini digunakan untuk membuat visualisasi data, dengan Seaborn untuk plot statistik dan Matplotlib untuk kustomisasi visualisasi.
- **OS:** Digunakan untuk berinteraksi dengan sistem operasi, termasuk pembuatan direktori output tempat hasil visualisasi disimpan.

#### b. Alur Kerja Analisis Data

Alur kerja analisis data dilakukan dalam empat tahap utama untuk memastikan konsistensi metodologis:

- **Persiapan Data:** Data didefinisikan langsung dalam skrip sebagai struktur data Python (dictionary) dan dikonversi menjadi DataFrame Pandas untuk menjaga kemandirian skrip dan menghindari kesalahan pembacaan file eksternal.
- **Pra-pemrosesan Data:** Untuk dataset dengan skala variabel berbeda (seperti data keuangan dan kurs valuta asing), dilakukan standarisasi menggunakan StandardScaler untuk menormalkan distribusi data. Pada data kurs, dilakukan feature engineering dengan menambahkan variabel 'Spread' (selisih kurs jual dan beli).
- **Implementasi K-Means:** Algoritma K-Means dari Scikit-learn dijalankan dengan jumlah kluster (k) yang telah ditentukan berdasarkan tujuan analisis (k=2 hingga k=3). Parameter random\_state digunakan untuk menghasilkan hasil klusterisasi yang konsisten.
- **Visualisasi dan Penyimpanan Hasil:** Hasil klusterisasi divisualisasikan menggunakan scatter plot dengan warna berbeda untuk tiap kluster dan label data ditambahkan sebagai anotasi. Visualisasi ini disimpan dalam format .png ke direktori output khusus.

#### c. Hasil Klusterisasi Data Kemiskinan

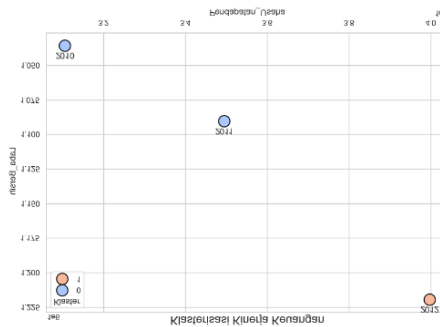


This work is licensed under a [Creative Commons Attribution-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/)



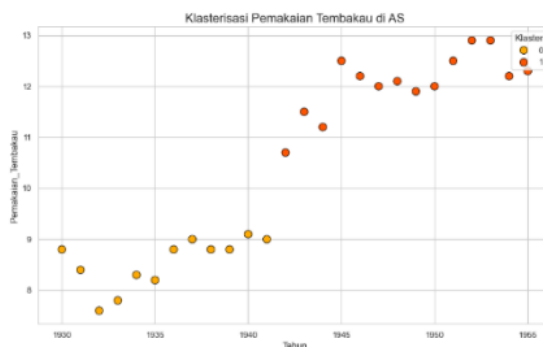
Dalam analisis data kemiskinan di Indonesia, K-Means mengelompokkan data tahunan ke dalam 3 klaster yang mencerminkan periode sejarah berbeda: tahun 1970-an dengan kemiskinan desa tinggi, tahun 1998 (krisis moneter) dengan lonjakan kemiskinan kota, dan tahun 2000-an dengan kemiskinan lebih terkendali. Temuan ini menunjukkan kemampuan K-Means dalam melakukan periodisasi historis berdasarkan karakteristik data.

**d. Hasil Klasterisasi Data Keuangan BUMN**



Meskipun dataset keuangan PT Angkasa Pura II hanya terdiri dari tiga titik data (tahun 2010, 2011, dan 2012), klasterisasi dengan  $k=2$  mampu memberikan wawasan yang berguna. Hasilnya secara efektif memisahkan tahun 2010 sebagai satu klaster tunggal, yang dapat diartikan sebagai kinerja baseline atau titik awal. Sementara itu, tahun 2011 dan 2012 dikelompokkan bersama dalam klaster kedua, yang mengindikasikan adanya kemiripan profil kinerja dan tren pertumbuhan pada kedua tahun tersebut.

**e. Hasil Klasterisasi Data Pemakaian Tembakau**



Pada data historis pemakaian tembakau di AS, K-Means dengan  $k=2$  berhasil menemukan titik perubahan signifikan. Visualisasi membagi data menjadi dua era: klaster pertama, "Era Pemakaian Rendah" sebelum 1942, dan klaster kedua, "Era Pemakaian Tinggi" setelahnya. Ini menunjukkan efektivitas K-Means dalam mendeteksi pergeseran tren meskipun hanya dengan satu variabel (univariat).







serta data cross-sectional, seperti anggaran pendidikan dan kurs valuta asing. Penggunaan standar fitur melalui StandardScaler terbukti penting dalam menghasilkan kluster akurat pada data dengan skala variabel berbeda. K-Means dapat diandalkan sebagai alat eksplorasi awal di berbagai disiplin ilmu, namun penelitian ini terbatas pada dataset kecil dan tidak membandingkan kinerja dengan algoritma clustering lain. Penelitian selanjutnya disarankan untuk menggunakan dataset lebih besar dan membandingkan K-Means dengan algoritma lain seperti Hierarchical Clustering atau DBSCAN untuk mengevaluasi keunggulan masing-masing metode

## DAFTAR RUJUKAN

- Bachtiar, A., Prasetyo, Y., & Nugroho, A. (2023). Perbandingan algoritma K-Means dan DBSCAN dalam pengelompokan data ekonomi daerah. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 10(2), 215–224.
- Bachtiar, A., Rizqi, N. I., & Mubarak, S. (2023). Komparasi Metode Clustering K-Means, Dbscan dan Hierarchical Untuk Analisis Penyakit Hepatitis C. *Journal Information System Development (ISD)*, 8(1), 27-32.
- Illah, S. N., Suarna, N., Ali, I., & Solihudin, D. (2025). K-Means Clustering Method to Make Credit Payment Grouping Efficient. *Journal of Artificial Intelligence and Engineering Applications (JAIEA)*, 4(2), 1074-1083.
- Irawan, D., Wijaya, G., & Warisaji, T. T. (2025). Penerapan Algoritma K-Means Clustering untuk Segmentasi Nasabah Bank. *BIOS: Jurnal Teknologi Informasi dan Rekayasa Komputer*, 6(1), 47-53.
- Irawan, R., Siregar, M., & Putra, D. (2023). Penerapan algoritma K-Means untuk segmentasi nasabah perbankan berbasis data transaksi. *Jurnal Sistem Informasi*, 19(1), 45–56.
- Kurniawan, H., & Defit, S. (2020). Data mining menggunakan metode k-means clustering untuk menentukan besaran uang kuliah tunggal. *Journal of Applied Computer Science and Technology*, 1(2), 80-89.
- Larasati, A., Maren, R., & Wulandari, R. (2021). Utilizing elbow method for text clustering optimization in analyzing social media marketing content of Indonesian e-commerce. *Jurnal Teknik Industri: Jurnal Keilmuan dan Aplikasi Teknik Industri*, 23(2), 111-120.
- Larasati, D., Wibowo, A., & Santoso, B. (2021). Penentuan jumlah kluster optimal menggunakan metode Elbow pada K-Means clustering. *Jurnal Ilmiah Informatika*, 6(2), 101–109.





- Putri, A. A. A., & Rahmah, S. A. (2024). Implementasi data mining dengan algoritma k-means clustering untuk analisis bisnis pada perusahaan asuransi. *Djtechno: Jurnal Teknologi Informasi*, 5(1), 139-152.
- Putri, R. A., & Rahmah, S. (2024). Segmentasi Nasabah Asuransi Menggunakan Algoritma K-Means Untuk Mendukung Keputusan Bisnis. *Jurnal Informatika dan Sistem Informasi*, 5(1), 33–42.
- Rahmah, S. A. (2024). Review Terbaru Tentang Klasterisasi Data Mining Menggunakan Metode K-Means: Tantangan dan Aplikasi. *Djtechno: Jurnal Teknologi Informasi*, 5(2), 297-303.
- Rasyid, S., Siswanto, S., & Sahriman, S. Clustering based on poverty indicator data using K-Means cluster with Density-Based Spatial Clustering of Application with Noise.
- Sano, A. V. D., & Nindito, H. (2016). Application of K-means algorithm for cluster analysis on poverty of provinces in Indonesia. *ComTech: Computer, Mathematics and Engineering Applications*, 7(2), 141-150.
- Sufairoh, I., Rani, A. C., Amalia, K., & Rolliawati, D. (2023). Perbandingan Hasil Analisis Clustering Metode K-Means, DBSCAN dan Hierarchical Pada Data Marketplace Electronic Phone. *JOINS (Journal of Information System)*, 8(1), 97-105.
- Sutramiani, N. P., Arthana, I., Aurelia, S., Fauzi, M., & Surya Darma, I. (2024). The Performance Comparison of DBSCAN and K-Means Clustering for MSMEs Grouping based on Asset Value and Turnover. *Journal of Information Systems Engineering & Business Intelligence*, 10(1).

