



Analisis Clustering K-Means dan K-Medoids terhadap Faktor-Faktor Kemiskinan untuk Karakterisasi Kemiskinan Kabupaten/Kota di Indonesia

Final Olimpiade Statistika SPSS 2024
by Data Dives

Maryesta Apriliani Sihombing
Farhan Akhtar Gymnasiar
Rachmania Azzahra Salsabila



Daftar Isi

1	Latar Belakang
2	Tujuan Penulisan
3	Metode Penelitian
4	Hasil dan Analisis
5	Kesimpulan dan Saran

Latar Belakang

Kemiskinan adalah fenomena multidimensional global yang dipengaruhi oleh berbagai faktor sosial, ekonomi, maupun demografi , dan telah menjadi isu krusial di banyak negara di dunia. Sebagai bagian dari komitmen global untuk meningkatkan kesejahteraan masyarakat, SDGs (Suistainable Development Goals) menempatkan pengentasan kemiskinan sebagai tujuan pertama dari 17 tujuan sasaran global yang harus dicapa. Hal ini menunjukkan bahwa kemiskinan masih menjadi tantangan utama di seluruh dunia (Leal Filho et al., 2021), di mana upaya pengentasan kemiskinan ini menjadi landasan bagi tercapainya tujuan pembangunan lainnya. Kemiskinan di Indonesia masih menjadi tantangan besar meskipun angka kemiskinan telah menurun dalam beberapa tahun terakhir. Berdasarkan data Badan Pusat Statistik (BPS) tahun 2024, persentase penduduk miskin tercatat sebesar 9,03%, tetapi angka ini tidak merata di seluruh wilayah. Hill (2021) menyatakan bahwa kesenjangan antarwilayah tetap menjadi masalah signifikan dan masih banyak daerah yang kesulitan menghadapi kemiskinan. Kondisi ini menunjukan bahwa diperlukan analisis yang lebih mendalam dan berdasarkan data untuk memahami faktor-faktor penyebab kemiskinan di setiap wilayah.

Tujuan Penulisan

Kemiskinan di Indonesia merupakan masalah kompleks yang dipengaruhi oleh berbagai faktor. Untuk memahami kemiskinan, diperlukan pembahasan yang lebih spesifik dan tepat sasaran, mengingat kondisi di berbagai kota dan kabupaten di Indonesia sangat beragam. Analisis yang mempertimbangkan karakteristik setiap daerah akan lebih mampu mengidentifikasi rekomendasi yang sesuai dengan kebutuhan dan tantangan spesifik masing-masing wilayah.

Tabel 1. Rangkuman hasil kajian literatur mengenai faktor kemiskinan di Indonesia

Refrensi	Proponents
<ul style="list-style-type: none">• Azis, dkk., (2023)• Purba, dkk., (2023)• Sinurat (2023)• Fitri (2020)• Prawoto & Basuki (2022)• Hendayanti & Nurhidayati (2020)	<ul style="list-style-type: none">• Jumlah penduduk, indeks pembangunan manusia, tingkat pengangguran, dan produk domestik bruto• Literasi, akses litrik rumah tangga, Produk Domestik Regional Bruto (PDRB)• Indeks pembangunan manusia• Tingkat pengangguran terbuka dan rata-rata pengeluaran per kapita sebulan untuk makanan• Pengeluaran daerah• Rata-rata lama sekolah dan harapan lama sekolah

Metodologi Penelitian



Data

Tabel 2. Variabel Penelitian

Variabel	Keterangan
X1	Persentase penduduk miskin dalam kabupaten/kota
X2	Rata-rata lama sekolah penduduk yang berusia lebih dari 15 tahun
X3	Pengeluaran perkapita (dalam satuan ribuan per orang per tahun)
X4	Indeks pembangunan manusia
X5	Usia harapan hidup
X6	Persentase rumah tangga yang memiliki akses sanitasi yang layak
X7	Persentase rumah tangga yang memiliki akses air minum layak
X8	Tingkat pengangguran terbuka
X9	Tingkat partisipasi angkatan kerja
X10	PDRB atas dasar harga konstan menurut pengeluaran (dalam satuan rupiah)

Data merupakan **data sekunder** yang berasal dari website kaggle.com. Setelah melakukan kajian literatur didapat bahwa 10 variabel tersebut terkait dengan **faktor kemiskinan**.

Selain 10 variabel tersebut, terdapat juga kolom bernama '**Klasifikasi Kemiskinan**' yang berisikan informasi status dari kabupaten/kota **miskin (1)** atau **tidak miskin (0)**

Metode

Metode Analisis Data:

Clustering

Clustering digunakan untuk dapat memahami karakteristik kabupaten/kota sehingga mengidentifikasi rekomendasi terkait kemiskinan yang sesuai dengan kebutuhan dan tantangan spesifik masing-masing klaster

Metode Clustering yang Digunakan:

K-Means dan

K-Medoids

Terdapat beberapa kondisi pada data yang ditemukan:

1. Memiliki variabel yang cukup banyak
2. Indikasi terdapat outlier
3. Variabel dengan pengukuran berbeda
4. Variabel merupakan data numerik

Setelah melakukan kajian pustaka pada jurnal/penelitian sebelumnya, akan digunakan kedua metode klastering ini

Tahapan Penelitian

1. Pemilihan Variabel

Problem: Variabel IPM

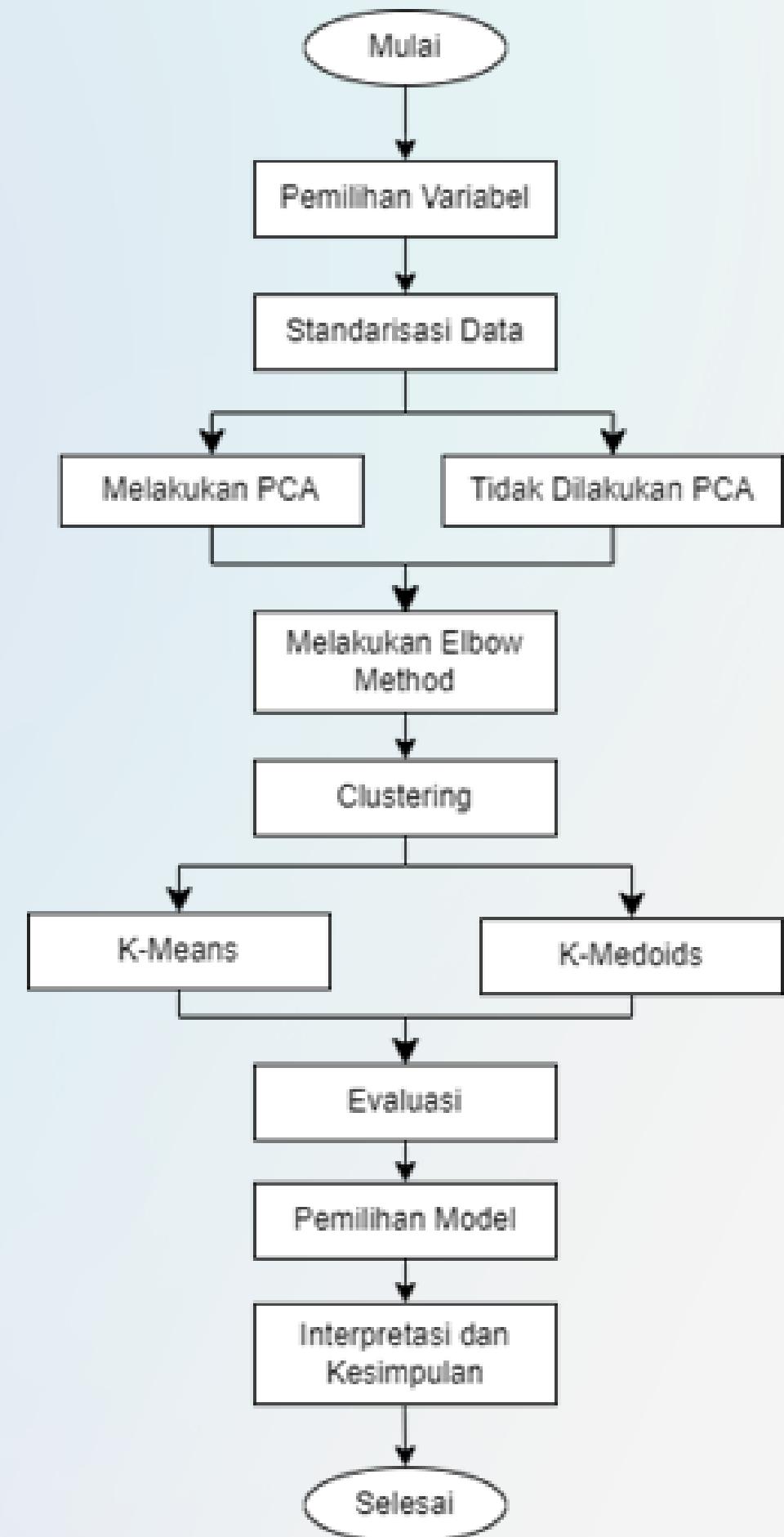
1. Merupakan hasil perhitungan beberapa variabel lain yang sudah tercakup dalam variabel lainnya.
2. Indikasi multikolinearitas. Nilai VIF sebesar 37.019905.

Solusi: IPM dikeluarkan dari variabel clustering

2. Standarisasi Data

Problem: Variabel memiliki skala pengukuran yang berbeda. Jika tidak diatasi, dapat menghasilkan klaster yang kurang akurat.

Solusi: Melakukan standarisasi data



Gambar 1. Flowchart Tahapan Penelitian

Tahapan Penelitian

3. Principal Component Analysis (PCA)

Problem: Terdapat 10 variabel. Dapat menyebabkan masalah "The Curse of Dimensionality".

Solusi: Reduksi dimensi dengan PCA. Akan dibandingkan klaster antara data yang sudah dan belum dilakukan PCA. Setelah dilakukan didapatkan 5 komponen utama yang mampu menjelaskan 85,64% variansi dari data.

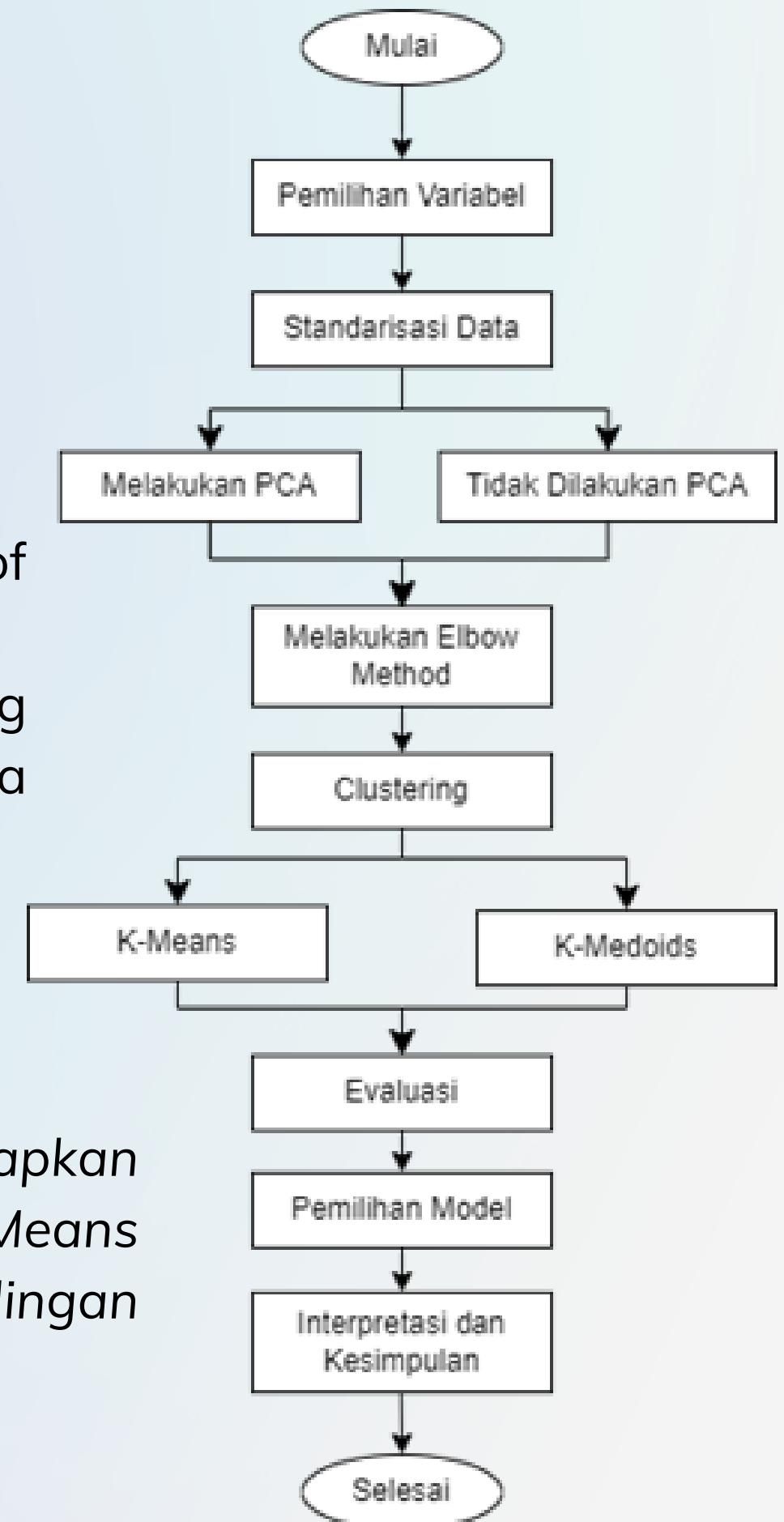
4. Elbow Method

Problem: Ingin diketahui jumlah klaster yang optimal

Solusi: Menentukan jumlah klaster yang optimal dengan Elbow Method yang diterapkan pada K-Means dan K-Medoids Clustering. Hasilnya, jumlah klaster optimal pada K-Means sebanyak 3, sementara pada K-Medoids sebanyak 4. akan dilakukan perbandingan performa antara jumlah klaster 3 dan 4 untuk kedua metode clustering tersebut.

5. Clustering

Digunakan metode K-Means Clustering dan K-Medoids Clustering



Gambar 1. Flowchart Tahapan Penelitian

Tahapan Penelitian

6. Evaluasi Hasil Clustering

Evaluasi akan menggunakan Silhouette Score dan Davies-Bouldin Index (DBI) sebagai metrik penilaian.

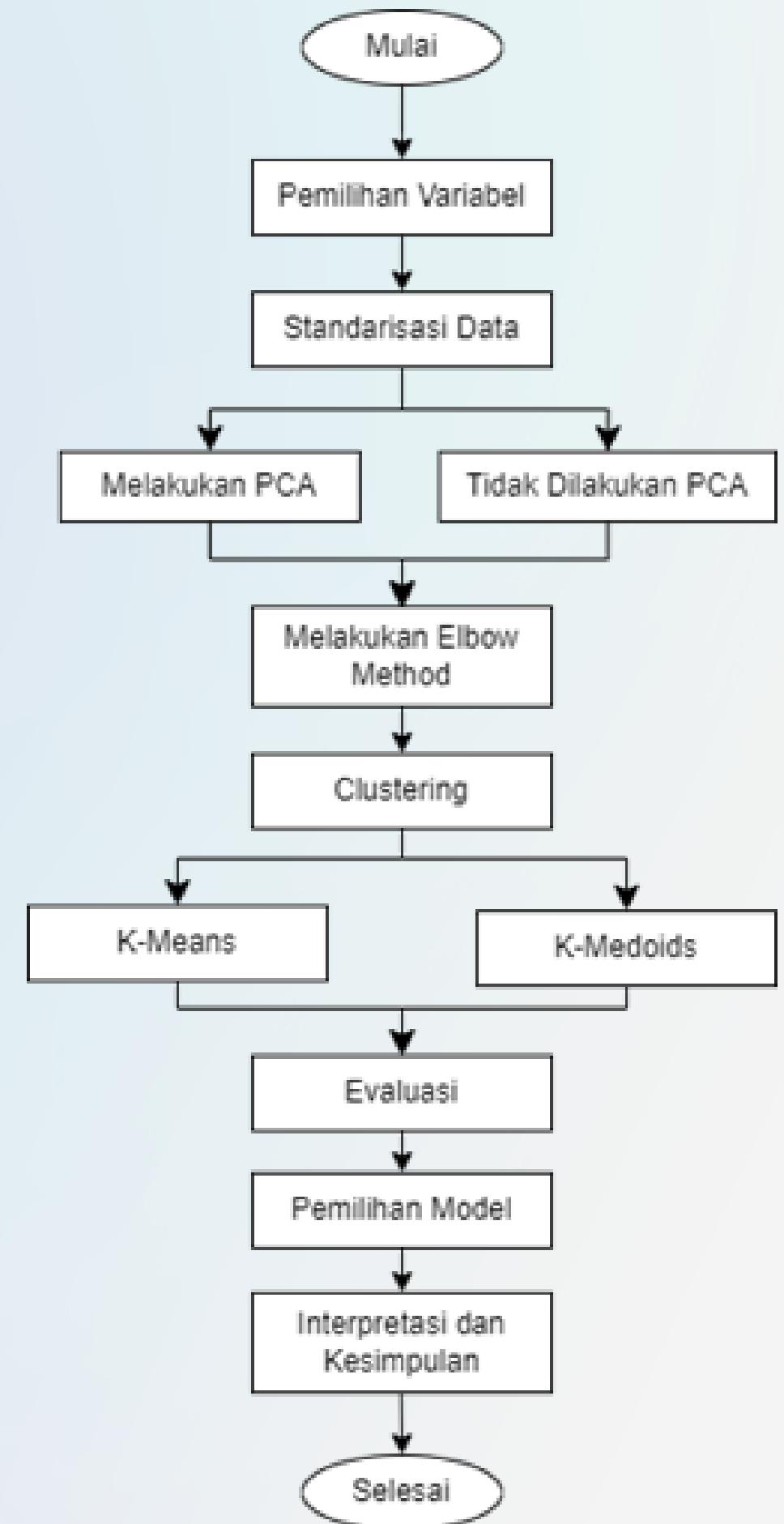
7. Pemilihan Clustering

Akan dipertimbangkan dengan melihat metrik evaluasi dan dengan urutan:

1. PCA atau tanpa PCA
2. K-Means Clustering atau K-Medoids Clustering
3. Jumlah klaster yang digunakan

8. Interpretasi dan Kesimpulan

Dilakukan karakterisasi untuk setiap klaster kemudian akan diberikan rekomendasi.



Gambar 1. Flowchart Tahapan Penelitian

Hasil dan Pembahasan



Hasil Clustering

PCA atau Tanpa PCA

Berdasarkan perbandingan rata-rata dari Silhouette Score dan Davies-Bouldin Index (DBI), **model clustering dengan PCA** menghasilkan performa yang **lebih baik** dibandingkan tanpa PCA.

Dipilih clustering dengan PCA

K-Means atau K-Medoids

K-Means memiliki performa yang lebih baik berdasarkan kedua metrik evaluasi.

Dipilih clustering K-Means dengan PCA

Jumlah Klaster

Dengan mempertimbangkan keterbacaan atau interpretasi dari hasil klastering, maka klaster dengan Silhouette Score lebih besar akan dipilih. Didapat Silhouette Score dengan jumlah 3 klaster lebih besar dibanding 4 klaster.

Clustering akhir yang dipilih adalah K-Means dengan 3 Klaster dan PCA

Tabel 7. Evaluasi dengan PCA dan Tanpa PCA

Clustering	Rata-Rata Silhouette Score		Rata-Rata DBI	
	Tanpa PCA	Dengan PCA	Tanpa PCA	Dengan PCA
K-Means	0.23256335	0.2880871	1.47826585	1.1269138
K-Medoids	0.1412517	0.2023536	1.9540752	1.9214633

Tabel 6. Evaluasi Tiap Model *Clustering*

Model		Silhouette Score		Davies-Bouldin Index (DBI)	
		3 Klaster	4 Klaster	3 Klaster	4 Klaster
K-Means	Tanpa PCA	0.2746607	0.190466	1.3116897	1.644842
	PCA	0.3272382	0.3055112	1.2323586	1.021469
K-Medoids	Tanpa PCA	0.1331244	0.149379	2.0925584	1.815592
	PCA	0.2027242	0.201983	2.1777466	1.66518

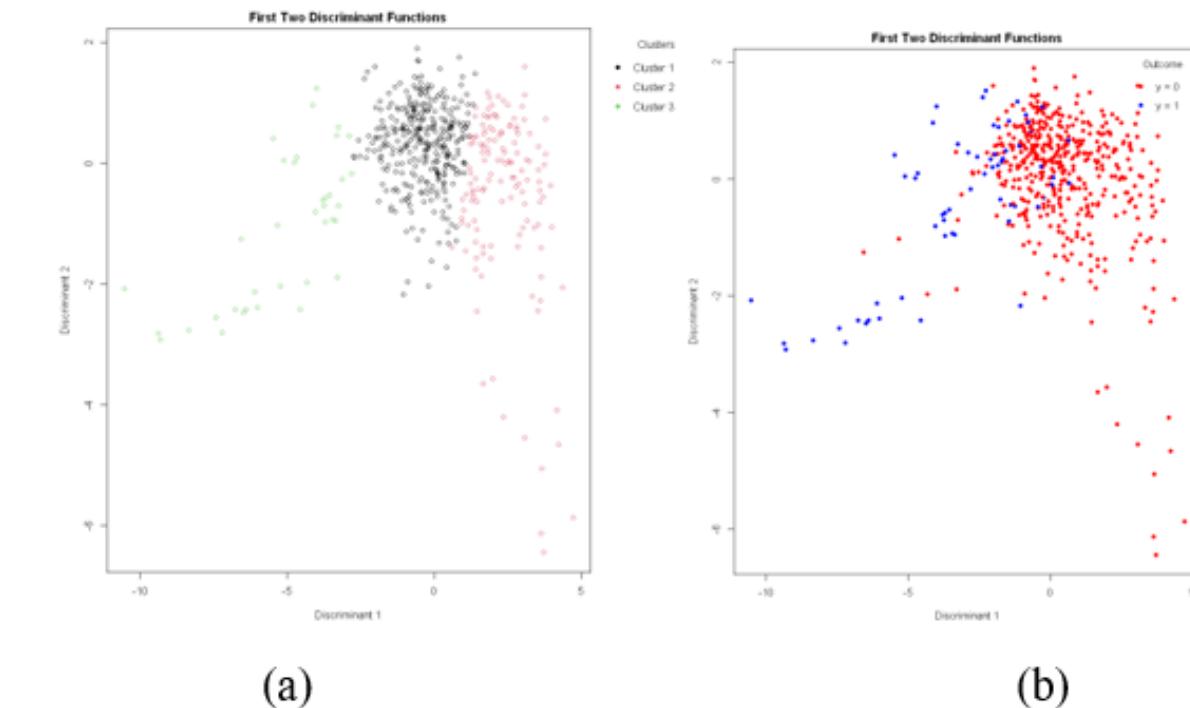
Interpretasi

Tabel 9. Informasi Hasil Clustering dengan Informasi Klasifikasi Kemiskinan

Klaster	Percentase Jumlah Kab/Kota dengan Status Miskin (%)	Jumlah Kab/Kota dengan Status Miskin	Jumlah Kab/Kota dengan Status Tidak Miskin	Jumlah Kab/Kota
1	81.1	30	7	37
2	0	0	137	137
3	9.41	32	308	340

Tabel 8. Rata-Rata untuk Setiap Variabel Penelitian

Klaster	X1	X2	X3	X5	X6	X7	X8	X9	X10
1	29.6	5.48	5938	65.3	35.8	54.9	1.96	81.3	1.84e6
2	7.46	9.98	13256	72.7	89.5	96.3	7.95	65.6	5.40e7
3	12.3	8.14	9621	68.9	76.8	83.9	4.23	69.7	1.12e7



(a)

(b)

Gambar 2. (a) Hasil dari *K-Means Clustering* dengan 3 Klaster dan *PCA* (b) Visualisasi dengan Informasi Klasifikasi Kemiskinan pada Data

Tabel 9. Informasi Hasil Clustering dengan Informasi Klasifikasi Kemiskinan

Klaster	Percentase Jumlah Kab/Kota dengan Status Miskin (%)	Jumlah Kab/Kota dengan Status Miskin	Jumlah Kab/Kota dengan Status Tidak Miskin	Jumlah Kab/Kota
1	81.1	30	7	37
2	0	0	137	137
3	9.41	32	308	340

Hasil Clustering

Klaster 1 (Kemiskinan Sangat Tinggi)

Penduduk memiliki tingkat kemiskinan yang tinggi (29.6%) dengan rendahnya rata-rata lama sekolah (5.48 tahun) dan pengeluaran per kapita yang rendah. Akses sanitasi dan air minum juga terbatas, dengan tingkat pengangguran yang rendah namun PDRB yang kecil.

Klaster 2 (Kemiskinan Sangat Rendah)

Tidak ada kabupaten/kota yang tergolong miskin, dengan pendidikan yang tinggi (9.98 tahun) dan pengeluaran per kapita tertinggi. Akses sanitasi dan air minum sangat baik, tetapi pengangguran cukup tinggi meski PDRB sangat besar.

Klaster 3 (Kemiskinan Menengah)

Tingkat kemiskinan sedang (12.3%) dengan rata-rata lama sekolah yang moderat (8.14 tahun) dan pengeluaran per kapita yang menengah. Akses sanitasi dan air minum cukup baik, dan tingkat pengangguran sedang dengan PDRB yang lebih rendah dari Klaster 2.

Tabel 8. Rata-Rata untuk Setiap Variabel Penelitian

Klaster	X1	X2	X3	X5	X6	X7	X8	X9	X10
1	29.6	5.48	5938	65.3	35.8	54.9	1.96	81.3	1.84e6
2	7.46	9.98	13256	72.7	89.5	96.3	7.95	65.6	5.40e7
3	12.3	8.14	9621	68.9	76.8	83.9	4.23	69.7	1.12e7

Tabel 2. Variabel Penelitian

Variabel	Keterangan
X1	Percentase penduduk miskin dalam kabupaten/kota
X2	Rata-rata lama sekolah penduduk yang berusia lebih dari 15 tahun
X3	Pengeluaran perkapita (dalam satuan ribuan per orang per tahun)
X4	Indeks pembangunan manusia
X5	Usia harapan hidup
X6	Percentase rumah tangga yang memiliki akses sanitasi yang layak
X7	Percentase rumah tangga yang memiliki akses air minum layak
X8	Tingkat pengangguran terbuka
X9	Tingkat partisipasi angkatan kerja
X10	PDRB atas dasar harga konstan menurut pengeluaran (dalam satuan rupiah)

Rekomendasi Kebijakan

Klaster 1 (Kemiskinan Sangat Tinggi)

- Tingkatkan akses pendidikan dan infrastruktur sanitasi.
- Ciptakan lapangan kerja untuk meningkatkan ekonomi lokal.

Klaster 2 (Kemiskinan Sangat Rendah)

- Atasi pengangguran dengan menciptakan lapangan kerja baru dan pelatihan keterampilan.
- Dorong investasi di sektor-sektor yang menciptakan pekerjaan.

Klaster 3 (Kemiskinan Menengah)

- Tingkatkan kualitas pendidikan dan infrastruktur dasar.
- Perkuat ekonomi lokal melalui pemberdayaan usaha kecil dan akses pasar.

Kesimpulan

Model clustering dengan kombinasi K-Means, 3 klaster dan PCA memberikan performa terbaik berdasarkan Silhouette Score dan Davies-Bouldin Index. Hasil analisis lebih lanjut menunjukkan bahwa klaster 1 merupakan klaster dengan tingkat kemiskinan yang sangat tinggi yang dapat terlihat pada persentase penduduk miskin yang paling tinggi diantara semua klaster sehingga diperlukan peningkatan program bantuan sosial dan pemberdayaan ekonomi sebagai upaya mengurangi tingkat kemiskinan yang sangat tinggi, klaster 2 memiliki tingkat kemiskinan yang sangat rendah dengan didukung oleh kualitas pendidikan yang baik, namun memiliki tingkat pengangguran yang tinggi sehingga diperlukan penanganan untuk mencegah peningkatan kemiskinan seperti memperluas akses lapangan kerja, dan klaster 3 adalah klaster dengan tingkat kemiskinan tertinggi kedua dengan pendidikan yang memadai, namun dengan tingkat pengangguran dalam kategori sedang sehingga diperlukan peningkatan program pendidikan dan pelatihan keterampilan.

Saran

Untuk penelitian selanjutnya, dapat digunakan variabel-variabel lain yang dapat lebih menggambarkan karakterisasi kemiskinan di Indonesia sehingga hasil yang didapat dapat memberikan berbagai macam insight. Selain itu, variasi penggunaan metode standarisasi, metode mencari jumlah klaster, metode clustering, maupun metrik evaluasi dapat dieksplorasi lebih jauh untuk didapatkan karakterisasi kabupaten/kota di Indonesia yang lebih baik.

Daftar Pustaka

- Adhitya, Prabawa, A., & Kencana, H. (2022). Analisis Pengaruh Pendidikan, Kesehatan, Sanitasi dan Rata-Rata Jumlah Anggota Keluarga Per Rumah Tangga terhadap Kemiskinan di Indonesia. *Ekonomis: Jurnal Ekonomi dan Bisnis*, 6(1), 288-295.
- Atira, & Sari, B. N. (2023). Penerapan Silhouette Coefficient, Elbow Method dan Gap Statistics untuk Penentuan Klaster Optimum dalam Pengelompokan Provinsi di Indonesia Berdasarkan Indeks Kebahagiaan. *Jurnal Ilmiah Wahana Pendidikan*, 9(17), 76-86.
- Azis, Sumertajaya, I. M., Purwaningsih, S. S., & Tjahjawati, S. S. (2023). Penentuan faktor kemiskinan Indonesia menggunakan regresi logistik. *Jurnal Matematika, Komputer, dan Statistik*, 6(1), 61–65.
- Buulolo, E. (2020). Data Mining Untuk Perguruan Tinggi (Edisi 1). Yogyakarta: Deepublish. [Online].
- Butsianto, & Mayangwulan. (2020). Penerapan Data Mining Untuk Prediksi Penjualan Mobil Menggunakan Metode K-Means Clustering. *Jurnal nasional Komputasi dan Teknol. Inf*, 3(3).
- Leal Filho, W., Lovren, V. O., Will, M., Salvia, A. L., & Frankenberger, F. (2021). Poverty: A central barrier to the implementation of the UN Sustainable Development Goals. *Ilmu & Kebijakan Lingkungan*, 125, 96-104.
- Hendayanti, N. P. N., & Nurhidayati, M. (2020). Regresi Logistik Biner dalam Penentuan Ketepatan Klasifikasi Tingkat Kedalaman Kemiskinan Provinsi-Provinsi di Indonesia. *Sainstek: Jurnal Sains dan Teknologi*, 12(2), 63-70.
- Hill, H. (2021). What's happened to poverty and inequality in Indonesia over half a century?. *Tinjauan Pembangunan Asia*, 38(1), 68-97.

Daftar Pustaka

- Khan, A. S. S., Fatekurohman, M., & Dewi, Y. S. (2023). Perbandingan Algoritma K-Medoids Dan K-Means Dalam Pengelompokan Kecamatan Berdasarkan Produksi Padi Dan Palawija Di Jember. *Jurnal Statistika dan Komputasi*, 2(2), 67-75.
- Ling, X., Tu, Q., Jin, M., Wang, W., Cui, Y., & Zhu, J. (2022, December). Research on Distributed Resource Aggregation Technology Based on Hierarchical Agglomerative Clustering Analysis. *Jurnal Fisika : Seri Konferensi* (Vol. 2399, No. 1, p. 012034). IOP Publishing.
- Mukhtar, S., & Saptono, A. (2019). Analisis pengaruh indeks pembangunan manusia dan tingkat pengangguran terbuka terhadap kemiskinan di Indonesia. *Ecoplan*, 2(2), 77-89.
- Muttaqin, M., & Anwar, K. (2023). Pengaruh Inflasi, Tingkat Pengangguran Terbuka, dan Tingkat Partisipasi Angkatan Kerja terhadap Tingkat Kemiskinan di Indonesia. *Jurnal Aplikasi Ilmu Ekonomi*, 2(2), 83-95.
- Prawoto, N., & Basuki, A. T. (2022). Factors Affecting Poverty in Indonesia: A Panel Data Approach. *Akses Kualitas Menuju Kesuksesan*, 23(186), 156-161.
- Purba, J. T., Budiono, S., Hariandja, E. S., & Pramono, R. (2023). Sustainability Strategy to Alleviate Poverty Through Education, Energy, GRDP, and Special Funds: Evidence from Indonesia. *International Journal of Sustainable Development & Planning*, 18(5).
- Sinurat, R. P. P. (2023). Analisis faktor-faktor penyebab kemiskinan sebagai upaya penanggulangan kemiskinan di Indonesia. *Jurnal Registratie*, 5(2), 87-103.
- Sopyan, Y., Lesmana, A. D., & Juliane, C. (2022). Analisis Algoritma K-Means dan Davies Bouldin Index dalam Mencari Klaster Terbaik Kasus Perceraian di Kabupaten Kuningan. *Building of Informatics, Technology and Science (BITS)*, 4(3), 1464-1470.
- Sugiharti, L., Purwono, R., Esquivias, M. A., & Jayanti, A. D. (2022). Poverty Dynamics in Indonesia: The Prevalence and Causes of Chronic Poverty. *Journal of Population & Social Studies*, 30.
- Wulandari, I., & Pratama, A. A. N. (2022). Analisis Pengaruh Dana ZIS (Zakat, Infak, Sedekah), Pertumbuhan Ekonomi, Angka Harapan Hidup, Rata-Rata Lama Sekolah, Dan Pengeluaran Perkapita Terhadap Tingkat Kemiskinan Di Indonesia Periode 2010-2021. *Jurnal Ilmiah Ekonomi Islam*, 8(3), 3301-3309.