



Pengelompokan Kualitas Pemuda Menurut Provinsi di Indonesia Tahun 2023 Menggunakan *Partition Clustering*

Faris Iqbal Maulana Susanto^{1,2}, Setia Pramana²

¹Program Studi Komputasi Statistik, Politeknik Statistika STIS

²Badan Pusat Statistik

¹222112045@stis.ac.id

²setia.pramana@stis.ac.id

Corresponding author email: 222112045@stis.ac.id

Abstract: The percentage of youth in Indonesia in the last decade has been in the range of 23-24 percent of the Indonesian population. Youth have an active role in national development. Therefore, youth development is the focus of the government because it is the key to accelerating development. This study aims to categorize provinces based on the condition of youth in health, education, and socio-economic aspects in 2023 so that it can assist the government in determining youth development policies effectively and inclusively. Based on the validation, the *k-medoids* algorithm was selected as the best clustering method with a cluster count of 4 and *Dunn*, *Davies-Bouldin*, and *Silhouette* index values of 0,323; 1,529; and 0,17, respectively. In general, the four clusters have quite different characteristics. Youth development in education, socio-economic, and health aspects by province still shows gaps. Therefore, equalization efforts tailored to the needs of provincial clusters are needed to achieve more equitable and inclusive progress.

Keywords: clustering, youth, *k-means*, *k-medoids*

Abstrak: Persentase pemuda di Indonesia dalam sedekade terakhir berada pada rentang 23-24 persen dari populasi Indonesia. Pemuda memiliki peranan aktif dalam pembangunan nasional. Oleh karenanya, pembangunan pemuda menjadi fokus pemerintah karena menjadi kunci percepatan pembangunan. Penelitian ini bertujuan untuk mengelompokkan provinsi berdasarkan kualitas pemuda pada aspek kesehatan, pendidikan, dan sosial ekonomi tahun 2023 sehingga dapat membantu pemerintah dalam menentukan kebijakan pembangunan pemuda secara efektif dan inklusif. Berdasarkan validasi yang dilakukan, algoritma *k-medoids* dipilih sebagai metode pengelompokan terbaik dengan jumlah klaster sebanyak 4 serta memiliki nilai indeks *Dunn*, *Davies-Bouldin*, dan *Silhouette* secara berturut-turut sebesar 0,323; 1,529; dan 0,17. Secara umum, keempat klaster memiliki karakteristik yang cukup berbeda. Pembangunan pemuda dalam aspek pendidikan, sosial ekonomi, dan kesehatan menurut provinsi masih menunjukkan adanya kesenjangan. Oleh karena itu, diperlukan upaya pemerataan yang disesuaikan dengan kebutuhan klaster provinsi untuk mencapai kemajuan yang lebih merata dan inklusif.

Kata kunci: clustering, pemuda, *k-means*, *k-medoids*

I. PENDAHULUAN

Menurut Undang-Undang Nomor 40 tahun 2009, pemuda adalah warga negara yang berusia 16 sampai 30 tahun. Pada tahun 2023, jumlah pemuda di Indonesia mencapai 64,16 juta jiwa atau 23,18 persen dari populasi, sedangkan dalam sedekade terakhir persentase pemuda berkisar pada angka 23 sampai 24 persen [1]. Jika dilihat menurut provinsi, persentase pemuda tertinggi berada di Papua dengan 26,64 persen, sedangkan persentase terendah berada di Jawa Timur dengan 21,09 persen pada tahun 2023. Persentase tersebut cukup besar sehingga memerlukan pemberdayaan pemuda yang optimal.

Pemuda merupakan aset bangsa, karena peran aktifnya dalam kontrol sosial, kekuatan moral, dan agen perubahan pada aspek pembangunan nasional [2]. Namun, perlu diketahui juga bahwa keberadaan pemuda menjadi beban yang berat karena berbagai macam kebutuhan yang harus dipenuhi untuk pemberdayaan pemuda [3]. Kebutuhan-kebutuhan tersebut antara lain: pendidikan, kelayakan hidup, dan lapangan pekerjaan. Oleh karenanya, pemberdayaan dan pembangunan pemuda menjadi salah satu fokus pemerintah karena sebagai kunci percepatan pembangunan [4].

Dalam mengukur pembangunan pemuda, Kementerian Pemuda dan Olahraga mengeluarkan Indeks Pembangunan Pemuda (IPP) tiap tahun [5]. Pengukuran ini menempatkan aspek pendidikan dan kesehatan pemuda pada lapisan pertama, yakni lapisan pembangunan individu. Hal tersebut didasarkan



bahwa kedua aspek tersebut adalah perwujudan dari pemenuhan hak kebutuhan dasar. Setelahnya, lapisan kedua yaitu lapisan pembangunan penghidupan dan kesejahteraan yang berisi aspek sosial ekonomi dasar pemuda, yakni akses terhadap lapangan dan kesempatan kerja. Dengan demikian, dalam merencanakan program pemberdayaan dan pembangunan pemuda ke depan, pemerintah perlu untuk memperhatikan kondisi terkini pemuda dari ketiga aspek tersebut pada setiap daerah. Penelitian ini diharapkan dapat mengelompokkan provinsi berdasarkan kualitas pemuda pada aspek kesehatan, pendidikan, dan sosial-ekonomi tahun 2023 sehingga dapat membantu pemerintah dalam menentukan kebijakan pembangunan pemuda secara efektif dan inklusif.

Penelitian ini merujuk pada penelitian yang dilakukan oleh Novaldi dan Wijayanto [6]. Penelitian tersebut membahas pengelompokan kualitas pemuda menurut provinsi di Indonesia tahun 2022 menggunakan metode hierarki dan *k-means*. Hasil terbaik diperoleh pada metode hierarki dengan jumlah kluster sebanyak dua. Selain itu, penelitian lain yang telah dilakukan adalah penelitian oleh Nuralifian dan Nooraeni [7] pada tahun 2021 mengenai pengelompokan indikator pemuda tahun 2021, diperoleh lima kluster menggunakan metode hierarki. Berdasarkan penelitian terdahulu, peneliti melakukan penelitian ini melalui pendekatan yang berbeda dari penelitian sebelumnya. Dari sisi metode, penelitian ini menggunakan metode *partition clustering*, yakni *k-means* dan *k-medoids* yang digunakan untuk analisis kluster. Pemilihan *partition clustering* sebagai metode pada penelitian ini dikarenakan metode ini sederhana dan mudah untuk diimplementasikan [8] serta cocok untuk dataset yang kecil [9]. Selain itu, dari sisi variabel yang digunakan, penelitian ini menggunakan beberapa variabel yang berbeda dibandingkan penelitian sebelumnya. Variabel yang digunakan pada penelitian ini berasal dari indikator-indikator kepemudaan pada tingkat provinsi yang bersumber dari Badan Pusat Statistik.

II. METODE PENELITIAN

2.1. Tinjauan Referensi

Analisis Kluster

Analisis kluster adalah metode numerik yang bertujuan untuk menemukan atau mengungkap kelompok pengamatan yang homogen dan terpisah dari kelompok lain [10]. Untuk mengelompokkan objek, diperlukan ukuran kemiripan yang umumnya menggunakan jarak. Jarak yang sering digunakan dalam data berskala numerik adalah jarak *Euclidean* [11]. Jarak ini dirumuskan seperti pada persamaan (1).

$$d = \sqrt{\sum_{j=1}^n (x_j - y_j)^2} \quad (1)$$

Sebelum melakukan analisis kluster, diperlukan beberapa asumsi yang harus dipenuhi [12], yakni kecukupan sampel dan non-multikolinieritas. Kecukupan sampel dapat diuji menggunakan uji *Kaiser-Meyer-Olkin* (KMO) [13]. Kecukupan sampel terpenuhi jika nilai KMO lebih besar dari 0,5. Untuk uji non-multikolinieritas, dapat diuji dengan melihat koefisien korelasi antarvariabel. Multikolinieritas terjadi jika terdapat variabel yang nilai korelasinya $> 0,8$ [14]. Adanya multikolinieritas pada analisis kluster dapat mempengaruhi hasil akhir pengelompokan [15].

Partition Clustering

Metode berbasis *partition* membagi objek-objek data menjadi k-bagian, yang disebut sebagai kluster [16]. Beberapa algoritma yang umum digunakan pada *partition clustering* adalah *k-means* dan *k-medoids* [17]. *K-means* menggunakan titik pusat (*centroid/means*) sebagai representasi kelompok, sedangkan *k-medoids* menggunakan objek yang paling representatif dalam kelompok



[18]. Keunggulan dari *k-means* adalah algoritmanya yang sederhana [19] dan cocok untuk data multidimensi, namun memiliki kelemahan yakni sensitif terhadap pencilan [20]. Di sisi lain, *k-medoids* mengatasi sejumlah kekurangan dari *k-means* [21]. Kedua algoritma ini juga sama-sama memiliki tantangan, yakni kebutuhan dalam penentuan jumlah kluster yang tepat [22].

Validasi Kluster

Validasi kluster dapat dilakukan melalui validasi internal. Terdapat banyak ukuran validasi internal [8]. Penelitian ini menggunakan indeks *Dunn*, indeks *Davies-Bouldin*, dan indeks *Silhouette*. Indeks *Dunn* mengukur seberapa baik kluster terpisah dan seberapa erat objek dalam setiap kluster, dengan menggunakan jarak terdekat antara objek di kluster berbeda untuk *separation* dan jarak terbesar dalam kluster yang sama untuk *compactness*, semakin besar nilainya maka semakin baik kluster yang dibentuk [23]. Indeks *Davies-Bouldin* (DB) dihitung dengan menentukan nilai tertinggi dari kemiripan antara setiap kluster dengan semua kluster lainnya, dan kemudian merata-ratakan nilai-nilai tertinggi tersebut sehingga jika nilainya lebih kecil menunjukkan kualitas *clustering* yang lebih baik [24]. Indeks *Silhouette* mengukur selisih berpasangan dari jarak *within-cluster* dan *between cluster*, sehingga semakin besar nilainya akan semakin baik [25].

2.2. Data dan Sumber Data

Data yang digunakan pada penelitian ini berasal dari Publikasi Badan Pusat Statistik, yaitu Statistik Pemuda 2023 [1]. Adapun variabel yang digunakan dalam penelitian ini berjumlah 10 variabel yang merepresentasikan aspek pendidikan, kesehatan, dan sosial ekonomi dengan rincian yang terdapat pada tabel 1.

Tabel 1. Rincian Variabel Penelitian

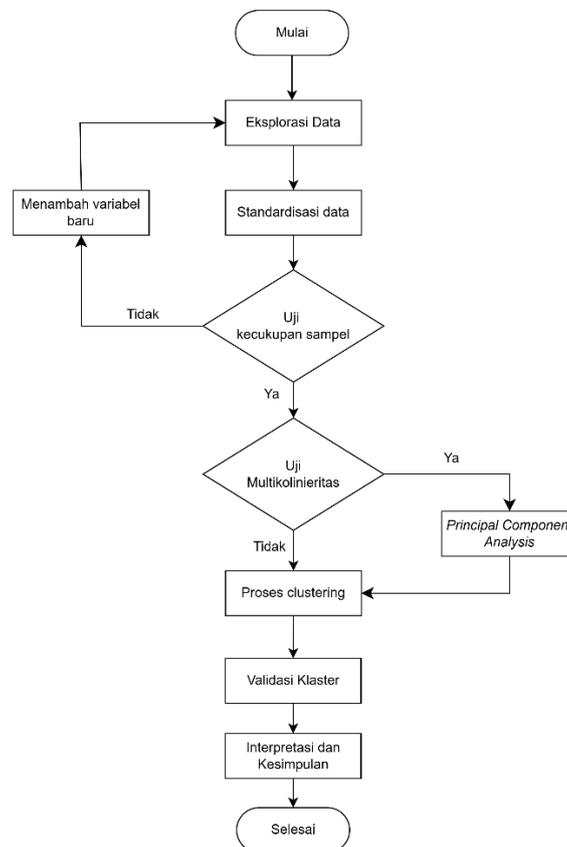
Aspek	Nama variabel	Satuan	Label
Pendidikan	Rata-rata lama sekolah pemuda	tahun	rls
	Tingkat partisipasi pemuda dalam pendidikan dan pelatihan formal dan nonformal	%	pelatihan
	Persentase pemuda yang memiliki akses internet	%	internet
Kesehatan	Angka kesakitan pemuda	%	kesakitan
	Persentase pemuda dengan jaminan kesehatan	%	jamkes
	Persentase pemuda dengan usia kawin pertama 16-18 tahun	%	kawin_dini
	Persentase pemuda perempuan yang melahirkan bayi dengan BBLR rendah	%	lahir_bblr
Sosial ekonomi	Tingkat pengangguran terbuka pemuda	%	tpt
	Persentase pemuda yang memiliki rekening tabungan	%	tabungan
	Persentase pemuda yang tinggal di rumah layak huni	%	layak_huni

2.3 Metode Analisis

Penelitian ini bersifat kuantitatif yang berfokus pada pengelompokan kualitas pemuda menurut provinsi di Indonesia tahun 2023 pada aspek pendidikan, kesehatan, dan ekonomi. Jumlah unit observasi pada penelitian ini adalah 34 provinsi. Penelitian ini menggunakan software *RStudio* untuk analisis kluster dan *QGIS* untuk memetakan persebaran kluster yang dibuat. Tahapan penelitian dapat dilihat pada gambar 1. Rincian dari tahapan penelitian dijelaskan sebagai berikut:

1. Eksplorasi data dari data yang telah dikumpulkan.

2. Setelah itu, data distandardisasi karena terdapat variabel yang memiliki satuan yang berbeda. Selain itu, proses ini juga memperkecil jarak antardata sehingga analisis kluster dapat dilakukan lebih tepat.
3. Uji kecukupan sampel dilakukan dengan uji *Kaiser-Meyer Olkin* (KMO). Apabila sampel tidak mencukupi, maka dapat menambah variabel, namun jika mencukupi, dilanjutkan ke proses berikutnya.
4. Uji multikolinieritas dengan melihat nilai korelasi antarvariabel. Apabila terjadi multikolinieritas, dilakukan *Principal Component Analysis*, namun jika tidak terjadi multikolinieritas, dilanjutkan ke tahap berikutnya.
5. Clustering menggunakan algoritma *k-means* dan *k-medoids*.
6. Validasi kluster menggunakan ukuran Indeks *Dunn*, *Davies-Bouldin*, dan *Silhouette*.
7. Interpretasi hasil kluster yang telah dibuat.



Gambar 1. Diagram alir penelitian

III. HASIL DAN PEMBAHASAN

3.1 Eksplorasi Data

Pada tabel 2, disajikan ringkasan tiap variabel. Pada aspek pendidikan, rata-rata lama sekolah pemuda terendah berada di Provinsi Papua, yakni 8,08 tahun. Dengan kata lain, rata-rata lama sekolah di Provinsi Papua hanya mencapai kelas dua sekolah menengah pertama. Sebaliknya, rata-rata lama sekolah pemuda tertinggi terdapat pada Provinsi DI Yogyakarta, yakni 12,61 tahun atau setara dengan lulus sekolah menengah atas.

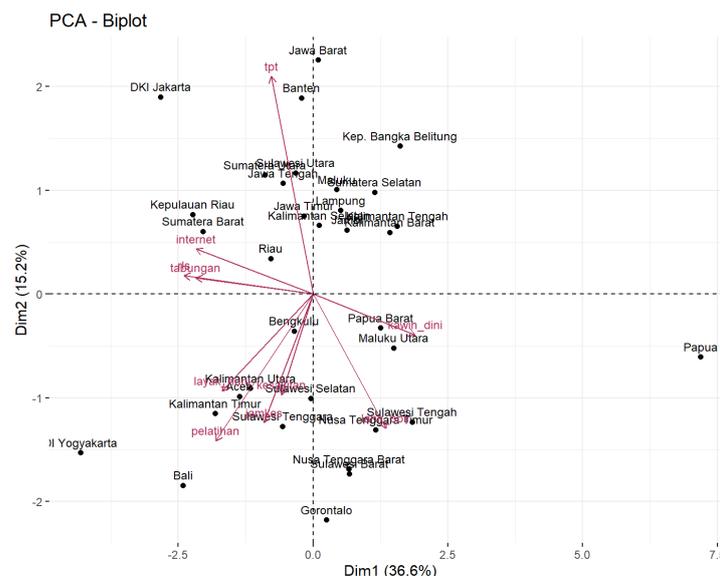


Tabel 2. Ringkasan variabel

Variabel	Minimum	Mean	Maksimum
rls	8,08	11,08	12,61
pelatihan	21,06	28,19	37,79
internet	37,26	91,78	99,01
kesakitan	3,11	6,17	13,33
jamkes	57,29	77,91	98,72
kawin_dini	7,34	21,17	31,60
lahir_bblr	9,03	13,47	20,17
tpt	5,37	11,29	19,64
tabungan	22,97	44,01	73,66
layak_huni	25,70	61,38	85,37

Pada aspek kesehatan, angka kesakitan pemuda tertinggi berada di Provinsi Nusa Tenggara Barat dengan 13,33 persen. Artinya, 13,33 persen pemuda di Provinsi NTB mengalami keluhan kesehatan sehingga aktivitas sehari-harinya terganggu. Di sisi lain, angka kesakitan pemuda terendah berada di Provinsi Papua dengan 3,11 persen. Persentase pemuda dengan usia kawin pertama di rentang 16-18 tahun tertinggi terdapat pada Provinsi Kalimantan Tengah yakni 31,6 persen. Sebaliknya, persentase terendah terdapat pada Provinsi Kepulauan Riau, yakni 7,34 persen.

Pada aspek sosial-ekonomi, Banten menjadi provinsi dengan tingkat pengangguran terbuka pemuda tertinggi yaitu sebesar 19,64 persen. Artinya, sekitar 20 dari 100 pemuda di Banten yang masuk dalam angkatan kerja, tidak terserap dalam pasar kerja. Nilai indikator TPT yang tinggi menunjukkan kurangnya potensi meraih keuntungan ekonomi. Sebaliknya, provinsi dengan tingkat pengangguran terbuka pemuda terendah adalah Sulawesi Barat yaitu sebesar 5,37 persen. Sementara itu, persentase pemuda yang tinggal di rumah layak huni terbesar terdapat di Provinsi DI Yogyakarta dengan 85,37 persen, sedangkan terendah terdapat di Provinsi Papua dengan 25,7 persen. Padahal, rumah yang tidak layak huni dapat mengakibatkan orang yang meninggalnya terkena risiko kesehatan.



Gambar 2. Biplot

Analisis biplot dilakukan dalam mereduksi data berdimensi tinggi menjadi dimensi yang lebih kecil (umumnya dua atau tiga dimensi) untuk memudahkan visualisasi. Metode yang digunakan dalam analisis ini adalah menggunakan dua komponen utama yang diperoleh pada metode *Principal Component Analysis* (PCA) [11]. Hal ini tentunya dapat menjadi gambaran mengenai hubungan

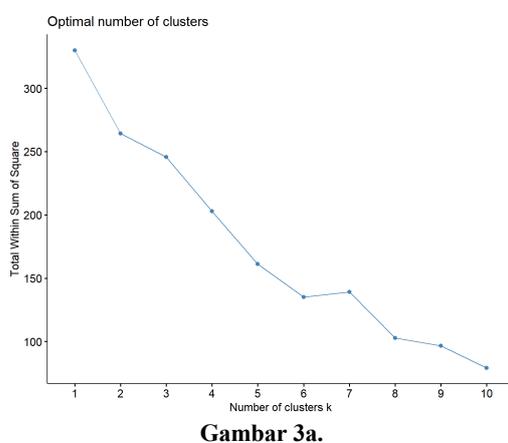
variabel dan letak antar objek. Analisis biplot disajikan pada gambar 2. Ukuran kesesuaian pada biplot tersebut hanya sebesar 51,8% sehingga hasil biplot belum cukup representatif, hal ini terjadi karena antar-atribut yang hampir saling independen satu sama lain (korelasi lemah). Pada biplot tersebut, nampak bahwa tiap variabel menunjukkan arah yang berbeda-beda, menandakan korelasi tiap variabel tersebut sangat lemah. Semakin panjang vektor suatu variabel menunjukkan bahwa semakin besar pula varians dari variabel tersebut, dan tampak bahwa atribut tingkat pengangguran terbuka memiliki keragaman yang besar dibanding atribut lainnya. Selain itu, letak beberapa titik provinsi jauh dibandingkan provinsi-provinsi yang lain, sehingga dimungkinkan dapat terindikasi sebagai pencilan.

3.2 Pengujian Asumsi

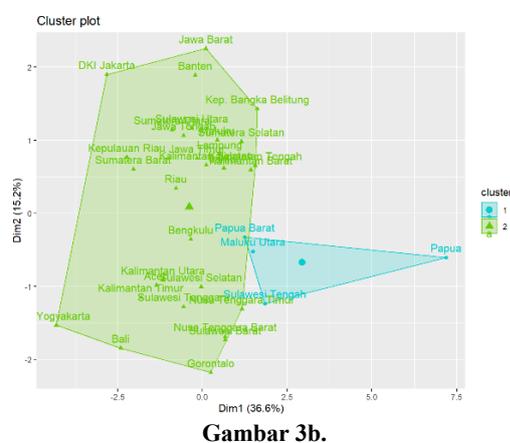
Uji asumsi pada *clustering* meliputi uji kecukupan sampel dan uji multikolinieritas. Uji kecukupan sampel menggunakan *Kaiser-Meyer-Olkin* (KMO) dihasilkan nilai KMO sebesar 0,647. Nilai tersebut lebih besar dari 0,5 yang menandakan bahwa sampel yang digunakan telah cukup untuk menggambarkan populasi yang ada. Setelah itu, dilakukan uji multikolinieritas dengan melihat korelasi untuk tiap variabel. Dari nilai korelasi yang diperoleh, nilai korelasi mutlak antarvariabel berada di rentang 0,04 sampai 0,66. Hal tersebut berarti tidak ada korelasi antarvariabel $> |0,8|$. Artinya, pada variabel-variabel yang digunakan tidak terjadi multikolinieritas sehingga dapat digunakan untuk tahap analisis kluster selanjutnya.

3.3 Proses Clustering

Sebelum melakukan pengelompokkan pada algoritma *k-means*, terlebih dahulu menentukan jumlah kluster optimal. Penentuan jumlah kluster optimal yang digunakan dapat menggunakan *elbow plot*. Jumlah kluster optimal pada *elbow plot* ditentukan pada titik di mana penurunan *within sum of square* (wss) mulai melambat secara signifikan, sehingga menciptakan siku pada grafik tersebut. Berdasarkan grafik *elbow plot* pada gambar 3a, nampak bahwa siku terletak pada $k = 2$, yang mana pada sebelum titik tersebut terjadi penurunan yang signifikan dan penurunan setelahnya mulai melambat. Oleh karenanya, dapat disimpulkan bahwa k optimal untuk algoritma *k-means* adalah 2 kluster. Berdasarkan gambar 3b, hasil pengelompokkan menggunakan *k-means* adalah kluster 1 beranggotakan 4 provinsi dan kluster 2 beranggotakan 30 provinsi.



Gambar 3a.

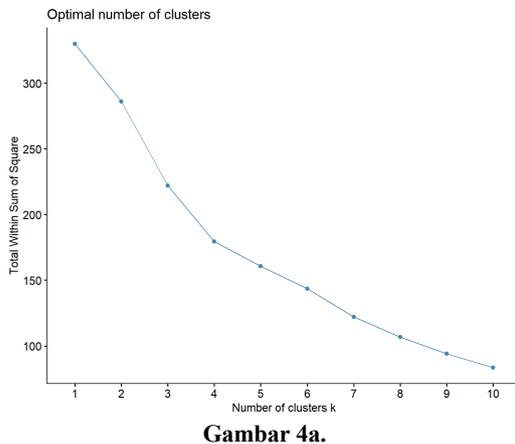


Gambar 3b.

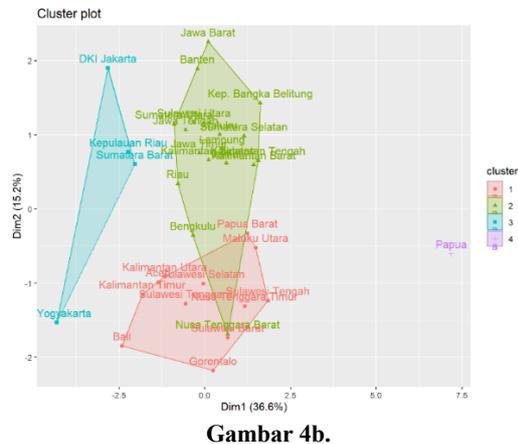
Gambar (2a). *Elbow plot* algoritma *k-means* (2b) Hasil clustering algoritma *k-means*

Serupa dengan *k-means*, pada algoritma *k-medoids*, juga menentukan jumlah kluster optimal terlebih dahulu menggunakan *elbow plot*. Berdasarkan grafik *elbow plot* pada gambar 4a, nampak bahwa siku terletak pada $k = 3$, yang mana pada sebelum titik tersebut terjadi penurunan yang signifikan dan

penurunan setelahnya mulai melambat. Oleh karenanya, dapat disimpulkan bahwa k optimal untuk algoritma k -medoids adalah 4 kluster.



Gambar 4a.



Gambar 4b.

Gambar (2a). Elbow plot algoritma k -medoids (2b) Hasil clustering menggunakan k -medoids

Setelah menentukan jumlah kluster optimal, selanjutnya adalah melakukan pengelompokkan dengan algoritma k -medoids dengan k optimal yang telah diperoleh. Hasil pengelompokkan menggunakan algoritma k -medoids ditunjukkan pada gambar 4b. Hasil pengelompokkan menggunakan k -medoids adalah kluster 1 beranggotakan 12 provinsi, kluster 2 beranggotakan 17 provinsi, kluster 3 beranggotakan 4 provinsi, serta kluster keempat beranggotakan 1 provinsi.

Hasil pengelompokkan kedua algoritma tersebut dapat dilihat pada gambar 3b untuk k -means dan 4b untuk k -medoids. Visualisasi tersebut mengacu pada biplot yang telah dianalisis sebelumnya pada tahap eksplorasi data. Visualisasi untuk algoritma k -means nampak lebih bagus dibandingkan dengan k -medoids karena hasil kluster lebih terpisah. Namun, ukuran kesesuaian dari visualisasi tersebut hanya 51,8% dari data yang sebenarnya, yang berarti banyak informasi yang hilang dari visualisasi tersebut [26]. Hal tersebut menjadi alasan bahwa pada data sebenarnya terdapat perbedaan jarak dan letak dari objek-objek tersebut.

3.4 Validasi Kluster

Setelah dilakukan pengelompokkan dengan kedua metode, kemudian dilakukan pemilihan metode terbaik melalui validasi kluster. Validasi kluster yang digunakan pada penelitian ini adalah indeks $dunn$, $davies-bouldin$ (DB), dan $silhouette$. Tabel 3 menjelaskan hasil validasi kluster untuk masing-masing metode. Dari tabel tersebut, nilai indeks $davies-bouldin$ terendah ditunjukkan oleh algoritma k -medoids dengan jumlah kluster sebanyak 4. Selain itu, algoritma k -medoids juga menghasilkan nilai indeks $dunn$ lebih tinggi dibanding k -means. Namun pada indeks $silhouette$, nilai tertinggi berada pada metode k -means. Dengan demikian, metode pengelompokkan $partitioning$ terbaik pada data kualitas pemuda menurut provinsi tahun 2023 adalah k -medoids.

Algoritma k -medoids cocok digunakan pada data dengan pencilan di dalamnya, sehingga menjadi alternatif algoritma k -means yang sensitif terhadap pencilan. Data yang digunakan dalam penelitian ini terdapat beberapa observasi yang diduga sebagai pencilan karena letaknya yang jauh dibandingkan dengan yang lain. Selain itu, algoritma k -medoids juga diterapkan pada jumlah observasi yang sedikit. Hal ini menjadi alasan k -means bukan metode $partition$ terbaik pada penelitian ini.

Tabel 3. Hasil validasi kluster

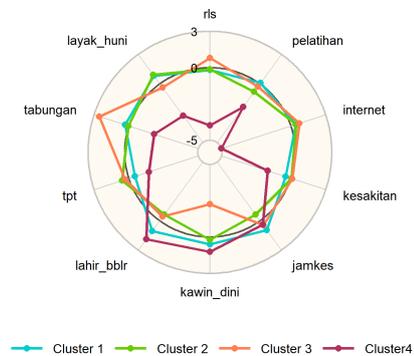
Metode	Jumlah k optimal	Ukuran validasi		
		<i>Dunn</i>	<i>DB</i>	<i>Silhouette</i>
<i>K-Means</i>	2	0,25318	1,6488	0,25
<i>K-Medoids</i>	4	0,32319	1,5288	0,17

3.5 Interpretasi

Berdasarkan hasil validasi kluster, diperoleh bahwa algoritma *k-medoids* memberikan hasil pengelompokan terbaik untuk data kualitas pemuda menurut provinsi tahun 2023. Jumlah kluster yang terbentuk adalah empat kluster. Rincian dari persebaran wilayah kluster yang terbentuk dapat dilihat pada peta tematik di gambar 5a.



Gambar 5a.



Gambar 5b.

Gambar (5a). Peta tematik (5b) *Radar plot*

Untuk mendapatkan informasi lebih detail mengenai kluster yang telah diperoleh, dilakukan profilisasi kluster. Agar lebih mudah dipahami, profilisasi kluster dapat disajikan melalui *radar plot* yang tersaji pada gambar 5b. Berdasarkan *radar plot*, dapat dijelaskan karakteristik dari kluster-kluster yang terbentuk:

1. Kluster pertama

Pada kluster pertama, aspek pendidikan sudah cukup baik. Hal ini ditunjukkan dengan rata-rata lama sekolah yang cukup tinggi. Persentase pemuda yang memiliki akses internet pun sudah baik serta tingkat partisipasi pemuda dalam pelatihan formal dan nonformal tertinggi dibandingkan kluster lain. Hal ini menunjukkan bahwa meskipun akses internet sudah baik, pemuda masih cenderung ikut dalam pelatihan formal dan nonformal walaupun dengan internet dapat mengeksplorasi secara otodidak. Pada aspek kesehatan, kluster pertama memiliki persentase pemuda yang kawin di bawah usia 19 tahun yang tinggi serta persentase pemudi yang melahirkan bayi dengan berat badan lahir rendah (BBLR) yang tinggi dibandingkan kluster lain. Kedua hal ini saling berkaitan, dimana pemudi yang kawin dini akan cenderung melahirkan bayi dengan BBLR rendah karena ketidaksiapannya dalam aspek psikologis dan ekonomi. Angka kesakitan pemuda pada kluster ini rendah dibandingkan dua kluster lainnya menunjukkan potensi yang besar pemuda untuk menjalankan aktivitas normal. Pada aspek sosial ekonomi, tingkat pengangguran terbuka pada kluster ini tergolong rendah dibandingkan kluster lain, memberikan indikasi bahwa tingkat penyerapan tenaga kerja sudah baik. Selain itu, sudah banyak pemuda yang memiliki rekening tabungan dan bertempat tinggal di rumah layak



huni. Hal ini menunjukkan adanya keterjaminan keuangan yang lebih baik serta indikasi keterjaminan kesehatan pemuda yang baik.

Oleh karenanya, pada klaster pertama dimungkinkan untuk meningkatkan kembali tingkat pendidikan formal bagi pemuda serta pendidikan dan akses terhadap kesehatan perlu ditingkatkan.

2. Klaster kedua

Pada klaster kedua, aspek pendidikan sudah cukup baik seperti klaster pertama. Hal ini ditunjukkan dengan rata-rata lama sekolah yang cukup tinggi. Persentase pemuda yang memiliki akses internet pun sudah baik, namun tingkat partisipasi pemuda dalam pelatihan formal dan nonformal rendah dibandingkan dua klaster lain. Hal ini menunjukkan bahwa karena akses internet yang sudah baik, pemuda cenderung dapat bereksplorasi terhadap pengetahuan secara otodidak. Pada aspek kesehatan, klaster kedua memiliki persentase pemuda yang kawin di bawah usia 19 tahun masih cukup tinggi serta persentase pemuda yang melahirkan bayi dengan BBLR yang rendah. Hal ini mungkin dikarenakan sudah matangnya pemuda saat kawin. Persentase pemuda yang memiliki jaminan kesehatan pada klaster ini merupakan yang terendah dibanding klaster lainnya. Pada aspek sosial ekonomi, tingkat pengangguran terbuka pada klaster ini tertinggi dibandingkan klaster lain, memberikan indikasi bahwa masih banyak pemuda yang menganggur, tidak terserap oleh pasar kerja. Selain itu, sudah banyak pemuda yang memiliki rekening tabungan dan bertempat tinggal di rumah layak huni sama seperti klaster pertama. Hal ini menunjukkan adanya keterjaminan keuangan yang lebih baik serta indikasi keterjaminan kesehatan pemuda yang baik.

Oleh karenanya, pada klaster kedua dimungkinkan untuk meningkatkan kembali tingkat pendidikan formal bagi pemuda agar dapat bersaing pada pasar kerja serta memperluas lapangan kerja. Pada aspek kesehatan, kemudahan untuk akses kesehatan perlu ditingkatkan karena angka kesakitan pemuda yang masih tinggi serta jaminan kesehatan juga perlu ditingkatkan untuk menghindari risiko ekonomi pada saat sakit.

3. Klaster ketiga

Klaster ketiga beranggotakan provinsi-provinsi dengan kualitas pemuda yang unggul dalam ketiga aspek. Pada aspek pendidikan, klaster ini jauh lebih baik dibandingkan klaster lainnya dibuktikan dengan rata-rata lama sekolah yang sangat tinggi. Persentase pemuda yang memiliki akses internet pun sudah baik, namun tingkat partisipasi pemuda dalam pelatihan formal dan nonformal tergolong tinggi. Pada aspek kesehatan, klaster ketiga memiliki persentase pemuda yang kawin di bawah usia 19 tahun terendah serta persentase pemuda yang melahirkan bayi dengan BBLR yang rendah. Hal tersebut mengindikasikan bahwa banyak pemuda di klaster ini yang lebih memilih untuk mencari pekerjaan dalam rangka mencukupi ekonomi dibandingkan kawin. Hal tersebut merupakan kondisi yang baik. Persentase pemuda yang memiliki jaminan kesehatan pada klaster ini tergolong tinggi, serta angka kesakitan pemuda yang tertinggi. Angka kesakitan pemuda yang tinggi menunjukkan bahwa banyak pemuda yang terganggu dalam melakukan aktivitas sehari-hari karena keluhan kesehatan. Pada aspek sosial ekonomi, tingkat pengangguran terbuka pada klaster ini masih tergolong tinggi, memberikan indikasi bahwa masih banyak pemuda yang menganggur, tidak terserap oleh pasar kerja. Selain itu, persentase pemuda yang memiliki rekening tabungan tertinggi dibandingkan klaster lain, namun masih banyak pemuda yang tinggal di rumah tidak layak huni. Hal ini



menunjukkan adanya keterjaminan keuangan yang lebih baik namun keterjaminan kesehatan pemuda yang kurang.

Oleh karenanya, pada klaster ketiga dimungkinkan untuk memperluas lapangan usaha agar mencukupi pasar tenaga kerja terutama para pemuda. Selain itu, perlu adanya sosialisasi pendidikan kesehatan dan rumah layak huni karena angka kesakitan pemuda yang masih tinggi serta masih banyak pemuda yang menempati rumah tidak layak huni.

4. Klaster keempat

Klaster keempat hanya beranggotakan Provinsi Papua. Di Papua, masih banyak sekali keteringgalan pemuda dibandingkan provinsi-provinsi lainnya pada ketiga aspek. Pada aspek pendidikan, baik rata-rata lama sekolah pemuda, tingkat partisipasi pelatihan formal dan nonformal pemuda, dan persentase pemuda yang mengakses internet menempati posisi terendah dan jauh tertinggal. Pada aspek kesehatan, persentase pemuda yang kawin dini serta pemuda yang melahirkan bayi yang BBLR juga tertinggi. Namun, angka kesakitan pemuda di Papua merupakan terendah dibandingkan klaster lain. Hal ini menunjukkan adanya potensi pemuda yang secara jasmani unggul dibandingkan yang lain. Pada aspek sosial ekonomi, tingkat pengangguran terbuka pada Papua justru terendah dibandingkan klaster lain. Hal tersebut dikarenakan banyak pemuda di wilayah pegunungan mayoritas bekerja sebagai pekerja keluarga tak dibayar. Di sisi lain, persentase pemuda yang tinggal di rumah layak huni juga masih jauh terendah dibandingkan klaster lain.

Hal yang paling dasar yang harus dibenahi pada Provinsi Papua adalah akses terhadap pendidikan. Dengan pendidikan, maka pengetahuan mengenai kesehatan dan aspek ekonomi akan membaik. Pembangunan infrastruktur pendidikan yang merata dan rumah layak huni perlu untuk ditekankan.

IV. KESIMPULAN

Berdasarkan hasil dan pembahasan pada bagian sebelumnya, diperoleh bahwa dari kedua algoritma yang diterapkan, algoritma pada metode *partition clustering* terbaik untuk data kualitas pemuda pada aspek pendidikan, kesehatan, dan ekonomi adalah algoritma *k-medoids* dengan jumlah klaster sebanyak 4. Secara umum, keempat klaster memiliki karakteristik yang cukup berbeda. Klaster pertama berisi provinsi dengan aspek pendidikan dan sosial yang cukup baik, namun aspek kesehatan masih kurang. Klaster kedua hampir sama dengan klaster pertama, namun memiliki aspek kesehatan yang lebih baik. Klaster ketiga berisi provinsi dengan kualitas pemuda yang unggul pada ketiga aspek. Klaster keempat tertinggal daripada klaster lainnya dari aspek yang paling mendasar yakni pendidikan. Pembangunan pemuda dalam aspek pendidikan, sosial ekonomi, dan kesehatan menurut provinsi masih menunjukkan adanya kesenjangan. Oleh karena itu, diperlukan upaya pemerataan yang disesuaikan dengan kebutuhan klaster provinsi untuk mencapai kemajuan yang lebih merata dan inklusif.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Politeknik Statistika STIS yang telah memberikan dukungan dana penelitian dalam menyelesaikan kajian ini.

REFERENSI

1. Badan Pusat Statistik Indonesia, *Statistik Pemuda Indonesia 2023*, vol. 21. Badan Pusat Statistik Indonesia, 2023. Accessed: Jun. 09, 2024. [Online]. Available:



- <https://www.bps.go.id/publication/2023/12/29/18781f394974f2cae5241318/statistik-pemuda-indonesia-2023.html>
2. Pemerintah Pusat Indonesia, *Undang-Undang Nomor 40 Tahun 2009 tentang Kepemudaan*. 2009. Accessed: Jun. 08, 2024. [Online]. Available: <https://peraturan.bpk.go.id/Details/38784/uu-no-40-tahun-2009>
 3. E. Sumantri, C. Darmawan, and Saefulloh, “Generasi dan Generasi Muda,” in *Pembinaan Generasi Muda*, 2nd ed., Tangerang Selatan: Universitas Terbuka, 2014, pp. 1–2.
 4. Pemerintah Indonesia, *Undang-Undang Nomor 40 Tahun 2009 tentang Kepemudaan*. Jakarta: Sekretariat Negara, 2009. Accessed: Jun. 08, 2024. [Online]. Available: <https://jdih.go.id/files/4/2009uu040.pdf>
 5. Kementerian Perencanaan Pembangunan Nasional (PPN)/Badan Perencanaan Pembangunan Nasional (Bappenas), *Indeks Pembangunan Pemuda 2021*. Jakarta: Direktorat Keluarga, Perempuan, Anak, Pemuda dan Olahraga Kementerian PPN/Bappenas, 2021. Accessed: Jun. 08, 2024. [Online]. Available: https://perpustakaan.bappenas.go.id/e-library/file_upload/koleksi/migrasi-data-publikasi/file/Unit_Kerja/Direktorat%20KPAPO/Indeks_Pembangunan_Pemuda_2021.pdf
 6. J. Novaldi and A. W. Wijayanto, “Analisis Cluster Kualitas Pemuda di Indonesia pada Tahun 2022 dengan Agglomerative Hierarchical dan K-Means,” *Komputika : Jurnal Sistem Komputer*, vol. 12, no. 2, pp. 91–99, Sep. 2023, doi: 10.34010/komputika.v12i2.10348.
 7. M. Z. Nuralifian and R. Nooraeni, “GROUPING PROVINCES IN INDONESIA BASED ON YOUTH DEVELOPMENT INDICATORS IN 2021,” *J-3P (Jurnal Pembangunan Pemberdayaan Pemerintahan)*, pp. 1–18, Jun. 2023, doi: 10.33701/j-3p.v8i1.3254.
 8. C. C. Aggarwal and C. K. Reddy, *Data Clustering Algorithms and Applications*, 1st ed. Chapman and Hall/CRC, 2014.
 9. G. Gandhi and R. Srivastava, “Review Paper: A Comparative Study on Partitioning Techniques of Clustering Algorithms,” *Int J Comput Appl*, vol. 87, no. 9, pp. 10–13, Feb. 2014, doi: 10.5120/15235-3770.
 10. B. Everitt and T. Hothorn, *An Introduction to Applied Multivariate Analysis with R*. New York, NY: Springer New York, 2011. doi: 10.1007/978-1-4419-9650-3.
 11. R. Johnson and D. Wichern, *Applied Multivariate Statistical Analysis*, 6th ed. Edinburgh: Pearson Education Limited, 2014.
 12. A. N. Fathia, R. Rahmawati, and T. Tarno, “ANALISIS KLASSTER KECAMATAN DI KABUPATEN SEMARANG BERDASARKAN POTENSI DESA MENGGUNAKAN METODE WARD DAN SINGLE LINKAGE,” *Jurnal Gaussian*, vol. 5, no. 4, pp. 801–810, Oct. 2016.
 13. A. C. Rencher and W. F. Christensen, *Methods of Multivariate Analysis*, 3rd ed. Provo: Wiley, 2012.
 14. M. Ridwan and S. Sunendiari, “Mendeteksi dan Mengatasi Multikolinieritas pada Data Penelitian Diabetes Melitus Wanita Suku Indian Tahun 2018,” in *Prosiding Statistika*, Feb. 2021, pp. 64–70.
 15. Sharma. Subhash, *Applied Multivariate Techniques*. 1995.
 16. Y. Zhang and E. Cheng, “An Optimized Method for Selection of the Initial Centers of K-Means Clustering,” in *Integrated Uncertainty in Knowledge Modelling and Decision Making*, Springer, 2013, pp. 149–156.
 17. S. Mousavi, F. Z. Boroujeni, and S. Aryanmehrn, “IMPROVING CUSTOMER CLUSTERING BY OPTIMAL SELECTION OF CLUSTER CENTROIDS IN K-MEANS AND K-MEDOIDS ALGORITHMS,” *J Theor Appl Inf Technol*, vol. 98, no. 18, pp. 3807–3814, Sep. 2020.
 18. D. Marlina, N. Iina, A. Fernando, and A. Ramadhan, “Implementasi Algoritma K-Medoids dan K-Means untuk Pengelompokkan Wilayah Sebaran Cacat pada Anak,” *Jurnal CoreIT*, vol. 4, no. 2, pp. 64–71, Dec. 2018.
 19. S. Pramana, B. Yuniarto, S. Mariyah, and R. Nooraeni, *Data Mining dengan R Konsep Serta Implementasi*, 1st ed. Bogor: IN MEDIA, 2018.



20. P. Govender and V. Sivakumar, “Application of k-means and hierarchical clustering techniques for analysis of air pollution: A review (1980–2019),” *Atmos Pollut Res*, vol. 11, no. 1, pp. 40–56, Jan. 2020, doi: 10.1016/j.apr.2019.09.009.
21. P. Arora, Deepali, and S. Varshney, “Analysis of K-Means and K-Medoids Algorithm For Big Data,” *Procedia Comput Sci*, vol. 78, pp. 507–512, 2016, doi: 10.1016/j.procs.2016.02.095.
22. J. Heidari, N. Daneshpour, and A. Zangeneh, “A novel K-means and K-medoids algorithms for clustering non-spherical-shape clusters non-sensitive to outliers,” *Pattern Recognit*, p. 110639, May 2024, doi: 10.1016/j.patcog.2024.110639.
23. J. C. Dunn†, “Well-Separated Clusters and Optimal Fuzzy Partitions,” *Journal of Cybernetics*, vol. 4, no. 1, pp. 95–104, Jan. 1974, doi: 10.1080/01969727408546059.
24. D. L. Davies and D. W. Bouldin, “A Cluster Separation Measure,” *IEEE Trans Pattern Anal Mach Intell*, vol. PAMI-1, no. 2, pp. 224–227, Apr. 1979, doi: 10.1109/TPAMI.1979.4766909.
25. P. J. Rousseeuw, “Silhouettes: A graphical aid to the interpretation and validation of cluster analysis,” *J Comput Appl Math*, vol. 20, pp. 53–65, Nov. 1987, doi: 10.1016/0377-0427(87)90125-7.
26. Z. A. Leleury and A. E. Wokanubun, “ANALISIS BILOT PADA PEMETAAN KARAKTERISTIK KEMISKINAN DI PROVINSI MALUKU,” *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, vol. 9, no. 1, pp. 21–31, Mar. 2015, doi: 10.30598/barekengvol9iss1pp21-31.