METODE K-MEANS CLUSTERING UNTUK MENGELOMPOKAN INDEKS PEMABANGUNAN MANUSIA (IPM) PROVINSI PAPUA

Kristia Yuliawan
Program Studi Informatika, STMIK Pesat Nabire
Jl. Poros Samabusa, Sanoba, Kec. Nabire, Kabupaten Nabire, Papua 98816
christianpesat@gmail.com

ABSTRAK

Kualitas sumber daya manusia merupakan bagian penting dalam menentukan keberhasilan suatu pembangunan. Pertumbuhan ekonomi yang tinggi tidak selalu menyelesaikan masalah kesejahteraan seperti kemiskinan dan taraf hidup masyarakat secara luas, sehingga keberhasilan pembangunan ini sering terlihat pada pencapaian kualitas Indeks Pembangunan Manusia (IPM). HDI memiliki tiga elemen penting yang membentuknya: kesehatan, pendidikan, dan populasi. Berdasarkan data Bappenas, menunjukkan bahwa IPM Papua berada di urutan 34 atau berada di urutan terbawah dari seluruh provinsi di Indonesia karena pembangunan yang tidak merata di tingkat kabupaten di Papua. Metode penelitian yang digunakan untuk clustering menggunakan algoritma K-Means dengan metode Elbow yang diimplementasikan menggunakan program python dengan aplikasi kode visual dan jupyter notebook. Penelitian ini bertujuan untuk membandingkan tingkat keakuratan klasifikasi IPM antar kabupaten di provinsi Papua. Hasil penelitian menunjukkan bahwa nilai klaster terbaik yang digunakan dalam penelitian ini adalah 2 dari jumlah klaster 29 kabupaten di provinsi Papua, di mana 12 kabupaten memiliki IPM rendah sedangkan 17 kabupaten memiliki IPM sedang (baik). Unsur-unsur penyusun IPM adalah pengeluaran per kapita, rata-rata sekolah, harapan sekolah, dan harapan hidup oleh karena itu, dengan bertambahnya nilai unsur-unsur yang membentuk IPM, IPM Provinsi Papua ke depan akan mengalami peningkatan yang signifikan

Kata kunci : *Clustering, Algoritma K-Means, Method Elbow, Python, Provinsi Papua, HDI.*

1. PENDAHULUAN

Indeks Pembangunan Manusia merupakan suatu indikator yang digunakan untuk mengukur tingkat pembangunan manusia di suatu wilayah. Indeks ini mempertimbangkan tiga dimensi dasar yaitu harapan hidup, pendidikan, dan standar hidup layak. Provinsi Papua merupakan salah satu daerah di Indonesia dengan IPM yang masih tergolong rendah, jauh tertinggal dibandingkan dengan daerah lain di Indonesia[1]. Amartya Sen menggambarkan indeks ini sebagai "pengukuran vulgar" oleh karena batasannya, namun indeks ini lebih berfokus pada hal-hal yang lebih sensitif dan berguna daripada hanya sekadar pendapatan per kapita yang selama ini di gunakan[2]. Dimensi kesehatan diukur dengan angka harapan hidup, dimensi pendidikan diukur dengan rata-rata lama sekolah dan angka melek huruf, sedangkan dimensi standar hidup diukur dengan kemampuan daya beli masyarakat. Berdasarkan nilai IPM, Indonesia termasuk ke dalam kategori menengah ke atas, demikian juga hal nya dengan seluruh daerah di Indonesia, namun Provinsi Papua masih tertinggal jauh[3]. Secara umum, IPM Indonesia mengalami peningkatan dari tahun ke tahun, meskipun demikian terdapat kesenjangan antar daerah yang signifikan di Indonesia[4]. Analisis lebih lanjut mengenai pengelompokan daerah berdasarkan IPM diperlukan untuk memahami masalah pembangunan manusia di Provinsi Papua.

Provinsi Papua memiliki IPM yang relatif rendah dibandingkan dengan provinsi lain di Indonesia, hal ini dipengaruhi oleh beberapa faktor, seperti akses layanan kesehatan dan pendidikan yang masih terbatas, serta kemiskinan yang masih tinggi[5], keadaan ini menunjukkan adanya tantangan dalam pembangunan manusia di Papua. oleh karena itu, diperlukan upaya-upaya strategis untuk meningkatkan IPM di Provinsi Papua, seperti pengembangan infrastruktur, peningkatan kualitas layanan dasar, dan pengentasan kemiskinan[6]. Pengelompokan IPM dapat membantu mengidentifikasi wilayah-wilayah yang memiliki karakteristik serupa sehingga intervensi pembangunan dapat dilakukan secara lebih tepat dan efektif sehingga informasi ini dapat digunakan untuk merumuskan kebijakan yang lebih tepat sasaran.

K-Means Clustering merupakan salah satu metode analisis klaster yang populer. Metode ini mengelompokkan data berdasarkan kesamaan karakteristik sehingga objek dalam satu klaster memiliki tingkat kemiripan yang tinggi[7]. Penerapan K-Means Clustering dapat membantu mengidentifikasi pola dan karakteristik IPM di Provinsi Papua[8]. Metode K-Means Clustering dapat digunakan untuk mengelompokkan data berdasarkan kemiripan karakteristik sehingga dapat membantu mengidentifikasi pola IPM di Provinsi Papua dan memberikan gambaran yang lebih jelas untuk perumusan kebijakan yang tepat[9].

Beberapa penelitian sebelumnya telah menggunakan metode K-Means Clustering untuk menganalisis IPM atau indikator pembangunan lainnya di Indonesia. Sebagai contoh, studi yang dilakukan oleh Dara Purnama Kessyi, dkk tentang metode clustering untuk menganalisis hubungan

dinamis antara pertumbuhan ekonomi dan ketimpangan pendapatan di Indonesia[10]. Selain itu, ada juga yang membandingkan efektivitas metode pembelajaran mesin seperti Random Forest, Support Vector Machine, dan AdaBoost dalam mengklasifikasi kabupaten/kota berdasarkan IPM yaitu penelitian yang dilakukan oleh Ressa Isnaini Arumnisaa dan Arie Wahyu Wijayanto[11]. Selain itu Penelitian lain di bidang kesehatan juga telah menerapkan K-Means Clustering untuk mengelompokkan data obat berdasarkan pola pemakaiannya. studi-studi sebelumnya telah mendemonstrasikan bahwa metode K-Means Clustering memiliki potensi yang besar untuk dimanfaatkan dalam menganalisis dan mengklasifikasikan berbagai jenis data terkait pembangunan, kesehatan, dan isu-isu sosial lainnya[12]

Penerapan metode K-Means Clustering untuk mengelompokkan IPM di Provinsi Papua dapat membantu mengidentifikasi karakteristik wilayah-wilayah dengan kemiripan tingkat pembangunan manusia. Melalui pengelompokan ini, dapat diperoleh gambaran mengenai distribusi dan pola IPM di Papua sehingga intervensi pembangunan dapat dilakukan secara lebih tepat dan efektif. Hasil pengelompokan IPM juga dapat menjadi informasi yang berharga bagi pembuat kebijakan untuk merumuskan strategi peningkatan pembangunan manusia yang sesuai dengan kebutuhan masing-masing klaster. Dengan demikian, metode K-Means Clustering memiliki potensi yang besar untuk membantu menganalisis dan mengurai permasalahan pembangunan manusia di Provinsi Papua[13].

Tujuan dari penelitian ini adalah untuk mengaplikasikan metode K-Means Clustering dalam mengelompokkan daerah-daerah di Provinsi Papua berdasarkan nilai Indeks Pembangunan Manusia (IPM). Dengan menggunakan teknik pengelompokan ini, diharapkan dapat diperoleh klasifikasi wilayah yang memiliki tingkat kemajuan pembangunan yang serupa, sehingga dapat membantu dalam pemetaan kondisi sosial ekonomi di berbagai daerah di Papua. Metode K-Means Clustering dipilih karena kemampuannya dalam mengidentifikasi pola dan struktur tersembunyi dalam data, yang pada gilirannya akan memberikan pemahaman lebih baik mengenai variasi IPM antar kabupaten atau kota di Papua. Selain itu, penelitian ini bertujuan untuk memberikan kontribusi terhadap pengambilan kebijakan pembangunan yang lebih tepat sasaran di Provinsi Papua. Dengan mengelompokkan wilayah-wilayah berdasarkan nilai IPM, hasil penelitian ini diharapkan dapat memberikan informasi yang berguna bagi pemerintah dan pemangku kepentingan lainnya dalam merancang program pembangunan yang lebih efisien dan terfokus pada daerah-daerah yang membutuhkan perhatian lebih. Melalui analisis ini, diharapkan dapat teridentifikasi wilayah yang memiliki kesenjangan pembangunan yang signifikan, sehingga dapat menjadi dasar dalam merancang intervensi kebijakan yang lebih sesuai dengan kondisi spesifik masing-masing kelompok wilayah.

2. TINJAUAN PUSTAKA

2.1. Indeks Pembangunan Manusia (IPM)

Indeks Pembangunan Manusia merupakan salah satu indikator untuk mengukur keberhasilan pembangunan suatu wilayah. IPM mencakup tiga dimensi dasar pembangunan manusia yaitu kesehatan, pendidikan, dan standar hidup yang layak[14]. Dimensi indikator kesehatan terdiri Angka Harapan Hidup, dimensi Pendidikan terdiri dari rata-rata Lama sekolah, dan dimensi Standar Hidup yang terdiri dari pengeluaran per Kapita Disesuaikan. Berdasarkan nilai IPM, suatu wilayah/negara dikelompokkan menjadi 4 kategori: kategori rendah (IPM < 60), menengah bawah, Menengah Atas, dan Tinggi[15].

2.2. Metode K-Means Clustering

K-Means Clustering merupakan salah satu metode analisis klaster yang populer dan sering digunakan. Tujuan dari K-Means Clustering adalah untuk mengelompokkan objek ke dalam K klaster berdasarkan kesamaan karakteristik[16]. Prinsip dari K-Means Clustering adalah meminimalkan jarak antara objek dan titik pusat (centroid) dari klaster yang bersangkutan. Kelebihan metode K-Means Clustering antara lain adalah kemampuannya untuk mengelompokkan data yang besar dengan cepat, kemudahan implementasi, serta kemampuannya untuk menangani data numerik. Di sisi lain, kekurangan K-Means Clustering adalah sensitif terhadap pemilihan titik pusat awal (initial centroids), serta kurang efektif untuk mengelompokkan data dengan bentuk non-convex[17]. Meskipun demikian, K-Means Clustering tetap menjadi salah satu metode clustering yang populer dan sering digunakan dalam berbagai bidang, termasuk untuk analisis pembangunan manusia[18].

2.3. Keterkaitan antara IPM dan Pembangunan Daerah

IPM merupakan salah satu indikator penting dalam menggambarkan capaian pembangunan manusia di suatu daerah. Pembangunan manusia tidak dapat dipisahkan dari pembangunan daerah karena keberhasilan pembangunan pada dasarnya ditujukan untuk meningkatkan kesejahteraan masyarakat. Pengelompokan IPM melalui K-Means Clustering dapat membantu mengidentifikasi karakteristik wilayah-wilayah dengan tingkat pembangunan manusia yang serupa. Hasil pengelompokan ini dapat dimanfaatkan dalam perumusan kebijakan pembangunan daerah yang lebih tepat sasaran dan sesuai dengan kondisi masing-masing kelompok [19].

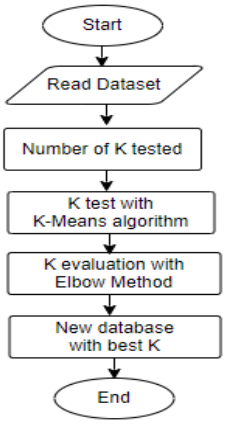
2.4. Peran K-Means Clustering dalam Analisis IPM

Penerapan K-Means Clustering untuk mengelompokkan IPM di Provinsi Papua memiliki beberapa manfaat misalnya mengidentifikasi pola dan

karakteristik IPM di Provinsi Papua secara lebih rinci, mengetahui wilayah-wilayah yang memiliki tingkat pembangunan manusia yang serupa danMembantu menentukan prioritas dan strategi pembangunan manusia yang sesuai dengan kebutuhan masing-masing klaster[7]. Dengan demikian, analisis K-Means Clustering dapat menjadi salah satu pendekatan yang bermanfaat untuk mengoptimalkan perumusan kebijakan pembangunan manusia di Provinsi Papua.

3. METODE PENELITIAN

Alur penelitian adalah desain sistem yang digunakan untuk melakukan pengelompokan dengan algoritma K_Means pada IPM Papua tahun 2020 dengan metode K-Means. Alur penelitian dalam penelitian ini dapat dijelaskan pada gambar di bawah ini.



Gambar 1. Alur Penelitian

3.1. Dataset

Dataset adalah kumpulan data yang terdiri dari lebih dari satu variabel yang terkait dengan topik tertentu. Misalnya, deep dataset diperlukan dalam pengelompokan HDI Provinsi Papua untuk mendapatkan informasi baru. Kumpulan data yang digunakan dalam penelitian ini bersifat publik, yaitu kumpulan data yang diambil dari *repositori publik yang* disepakati oleh berbagai ahli. Kumpulan data dalam penelitian ini diambil dari Koggle, platform terbaik di komunitas ilmu data.

3.2. Jumlah K yang diuji

Nomor database kluster awal adalah input awal dari kluster yang diperlukan untuk menghasilkan output kluster akhir. Dengan menggunakan algoritma K-Means, nilai K sebagai titik awal dipilih secara acak. Pada tahap ini, kita menentukan variabel yang ingin kita lakukan pengelompokan. Langkah selanjutnya adalah melakukan penskalaan data, yaitu, menormalkan ukuran data dengan perbedaan jarak yang signifikan antar variabel, yang dapat membuat plot tidak muncul dengan benar.

3.3. Pengujian K dengan Algoritma K-Means.

K-Means adalah metode pengelompokan data non-hierarkis yang mencoba mempartisi data yang ada menjadi satu atau lebih kluster atau grup. Metode ini membagi ke dalam kelompok atau kelompok sehingga data dengan karakteristik yang sama (kesamaan kelas High Intra) dikelompokkan ke dalam kelompok yang sama, dan yang memiliki karakteristik berbeda (Kesamaan antar kelas Law) dikelompokkan ke dalam kelompok lain. Proses pengelompokan dimulai dengan mengidentifikasi data yang akan dikelompokkan, Xij (i=1,...,n; j=1,...,m), di mana n adalah jumlah data yang akan dikelompokkan dan m adalah jumlah variabel. Pada awal iterasi, pusat setiap cluster diatur secara independen (sewenang-wenang), Ckj (k=1,...,k; j=1,...,m). Kemudian menghitung jarak antara setiap data dengan masing-masing pusat klaster. Untuk menghitung jarak ke data ke-i (xi) di pusat k-cluster (ck), diberi nama (dik), dapat digunakan rumus Euclidean, seperti pada persamaan (1), yaitu:

$$d_{ik} = \sqrt{\sum_{j=1}^{m} (x_{ij} - c_{ij})^2} \tag{1}$$

Sebuah data akan menjadi anggota k-cluster jika jarak ke pusat k-cluster adalah nilai terkecil dibandingkan dengan jarak ke pusat kluster lainnya. Ini dapat dihitung menggunakan persamaan (2). Selanjutnya, kelompokkan data yang merupakan anggota dari setiap kluster

$$Min \sum_{k=1}^{k} d_{ik} = \sqrt{\sum_{j=1}^{m} (x_{ij} - c_{ij})^2} \tag{2}$$

Nilai pusat klaster baru dapat dihitung dengan mencari nilai rata-rata data yang merupakan anggota klaster, menggunakan rumus pada persamaan (3):

$$c_{ij} = \frac{\sum_{i=1}^{p} x_{ij}}{p} \tag{3}$$

Deskripsi: Dimana kluster ke-k $x_{ij} \in$, p = jumlah anggota kluster ke-k.

3.4. Evaluasi K dengan metode elbow

Metode Elbow membandingkan nilai atau persentase beberapa k yang diuji dan membentuk elbow pada suatu titik. Dalam metode elbow, kita mendapatkan kluster terbaik dengan melihat hasil persentase dalam perbandingan antara jumlah gugus yang membentuk sudut untuk suatu sudut. Misalnya, kluster terbaik diperoleh jika kluster pertama dan kedua pergi ke sudut dalam grafik. Perbandingan antara jumlah gugus dilakukan dengan menghitung SEE (Sum of Square Error) dari setiap kluster, sedangkan rumusnya seperti persamaan (4):

$$SSE = \sum_{k=1}^{k} \sum_{xi} | xi - c^2 \quad (4)$$

Deskripsi: K = c-cluster Xi = jarak data objek ke-i Ck = pusat kluster ke-i

3.5. Menggunakan K terbaik untuk database baru

Langkah terakhir setelah cluster terbaru dibentuk untuk digunakan pada dataset di mana atribut dapat berasal dari atribut yang berbeda sehingga dataset sudah memiliki label

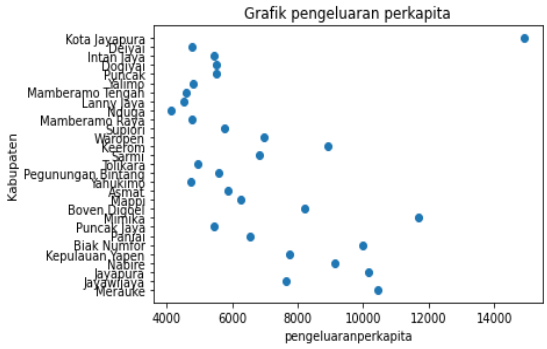
4. HASIL DAN PEMBAHASAN

4.1. Sumber Dataset

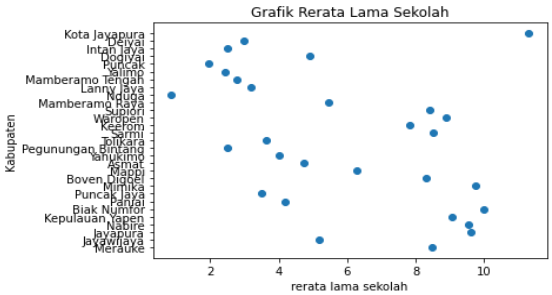
Dataset yang digunakan dalam penelitian ini diambil dari situs kaggele, mengakomodasi aktivis ilmu data yang ingin mempelajari lebih lanjut tentang pembelajaran mesin dan ilmu terkait lainnya. Adapun perintah untuk menampilkan dataset menggunakan visual studio code dan jupyter notebook dengan pustaka pandas dan mathpoltlib adalah sebagai berikut :

Gambar 2. Dataset

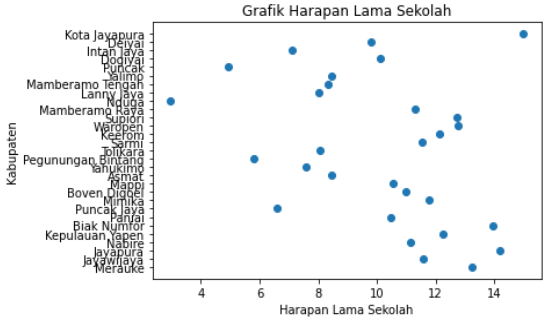
Setelah membuka file dataset, kemudian divisualisasikan data IPM Provinsi Papua tahun 2020, berdasarkan data pengeluaran per kapita, rata-rata lama sekolah, harapan sekolah, dan harapan hidup.



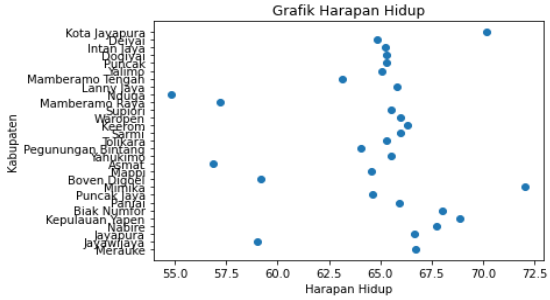
Gambar 3. Grafik Distribusi Pengeluaran per kapita



Gambar 4. Grafik Distribusi Rata-rata Lama Sekolah



Gambar 5. Grafik Ekspektasi Sekolah Lama



Gambar 6. Bagan Harapan Hidup

4.2. Jumlah K yang diuji

Langkah selanjutnya adalah menentukan jumlah pengelompokan untuk data pelatihan. Untuk data pelatihan, elemen yang digunakan untuk menyusun IPM adalah atribut pengeluaran per kapita, rata-rata lama sekolah, harapan sekolah, dan harapan hidup. Hasil data pengelompokan yang akan diuji dapat dilihat pada gambar di bawah ini.

SLICING

	0	1	2	3
0	10430	8.49	13.24	66.71
1	7637	5.17	11.58	58.99
2	10160	9.6	14.17	66.66
3	9143	9.53	11.14	67.72
4	7739	9.07	12.24	68.85
5	9969	10	13.94	68
6	6535	4.2	10.47	65.94
7	5459	3.51	6.59	64.65
8	11700	9.76	11.77	72.06
9	8211	8.32	10.99	59.16
10	6268	6.29	10.53	64.56
11	5882	4.74	8.47	56.88
12	4737	4.01	7.59	65.52
13	5578	2.49	5.79	64.08
14	4946	3.62	8.04	65.3
15	6814	8.52	11.55	66
16	8918	7.83	12.14	66.35
17	6978	8.87	12.77	65.99

Gambar 7. Data Pelatihan

Untuk menghasilkan hasil terbaik dari pengelompokan HDI Provinsi Papua pada tahun 2020, kita perlu mengubah bingkai data menjadi bentuk array sehingga jarak antar variabel terlihat. Dalam penelitian ini, jarak yang menjadi patokan adalah nilai 0 dan 1 yang disebut penskalaan atau normalisasi data dengan menggunakan skala 0 dan 1 sehingga

pengelompokan dapat dilakukan secara optimal. Hasilnya berupa data array, seperti yang ditunjukkan di bawah ini.

```
array([[0.58372718, 0.73110048, 0.85465116, 0.68967517],
       [0.32490038, 0.41339713, 0.71677741, 0.24187935],
       [0.55870633, 0.83732057, 0.93189369, 0.68677494],
       [0.46446113, 0.83062201, 0.68023256, 0.74825986],
       [0.3343527 , 0.78660287, 0.77159468, 0.8138051 ],
       [0.54100639, 0.87559809, 0.9127907 , 0.76450116],
       [0.22277824, 0.32057416, 0.62458472, 0.6450116 ],
       [0.12306552, 0.25454545, 0.30232558, 0.57018561],
       [0.70141785, 0.85263158, 0.73255814, 1.        ],
       [0.37809286, 0.71483254, 0.66777409, 0.25174014],
       [0.1980354 , 0.52057416, 0.62956811, 0.5649652 ],
       [0.16226485, 0.3722488 , 0.45847176, 0.11948956],
       [0.05615791, 0.30239234, 0.38538206, 0.62064965],
       [0.13409323, 0.1569378 , 0.2358804 , 0.53712297],
       [0.0755259 , 0.26507177, 0.42275748, 0.60788863],
       [0.24863312, 0.73397129, 0.71428571, 0.64849188],
       [0.44361042, 0.66794258, 0.76328904, 0.6687935 ],
       [0.26383097, 0.76746411, 0.81561462, 0.64791183],
       [0.15179316, 0.7215311 , 0.81146179, 0.6212297 ],
       [0.05782597, 0.44114833, 0.69352159, 0.13689095],
       [0.        , 0.        , 0.        , 0.        ],
       [0.03577055, 0.22296651, 0.42026578, 0.6363109 ],
       [0.04429617, 0.184689  , 0.44684385, 0.48259861],
```

Gambar 8. Hasil Penskalaan Data

4.3. Pengujian K dengan Algortima K-Means

Pada bagian ini, kita dapat menggunakan algoritma K-Means Clustering menggunakan pustaka Scukit Learn dengan menentukan jumlah klaster. Hasil dari jumlah pengelompokan yang digunakan dapat dilihat pada gambar di bawah ini.

```
KMeans
KMeans(n_clusters=4)
```

Gambar 9. Jumlah Algoritma Pengklusteran K-Means

Hasil pelabelan nilai kluster dengan nilai 4 dapat dilihat pada gambar di bawah ini.

```
array([2, 0, 2, 2, 2, 2, 1, 1, 2, 0, 0, 0, 1, 1, 1, 2, 2, 2, 2, 0, 3, 1,
       1, 1, 1, 1, 1, 2])
```

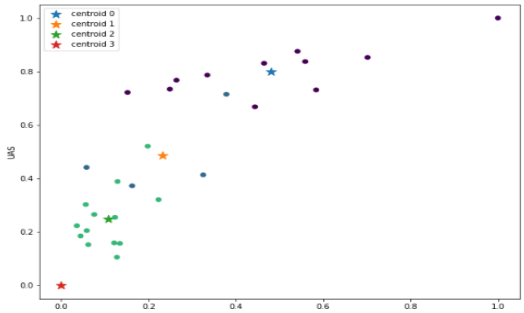
Gambar 10. Label Titik Data

Setelah diberi nilai 4 untuk setiap titik pusat dalam klaster, maka akan menghasilkan pengelompokan dengan titik pusat di setiap klaster seperti gambar di bawah ini.

```
array([[0.48104902, 0.80043497, 0.81712474, 0.74351403],
       [0.23077101, 0.4854067 , 0.63413621, 0.1875    ],
       [0.10674137, 0.24902466, 0.43074367, 0.5896841 ],
       [0.        , 0.        , 0.        , 0.        ]])
```

Gambar 11. Hasil Nilai Pengklusteran

Visualisasi data pengelompokan dari data klaster dalam penelitian disertai dengan warna yang berbeda sehingga lebih mudah bagi kita dalam menganalisisnya dapat dilihat pada gambar di bawah ini.



Gambar 12. Visualisasi Pengklusteran Data

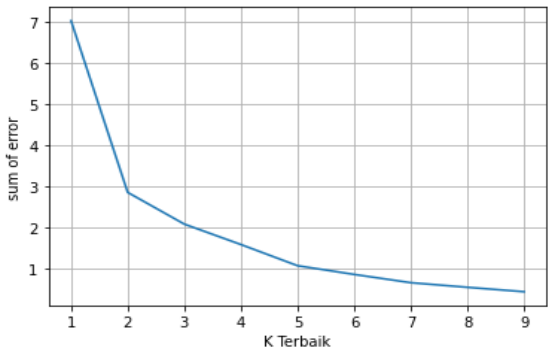
Dataset akan mengalami perubahan dari dataset awal di mana atribut klaster terbentuk pada dataset HDI Provinsi Papua pada tahun 2020.

	Kabupaten	pengeluaranperkapita	reratalamasekolah	harapanlamasekolah	angkaharapanhidup	cluster
0	Merauke	10430.0	8.49	13.24	66.71	2
1	Jayawijaya	7637.0	5.17	11.58	58.99	0
2	Jayapura	10160.0	9.60	14.17	66.66	2
3	Nabire	9143.0	9.53	11.14	67.72	2
4	Kepulauan Yapen	7739.0	9.07	12.24	68.85	2
5	Biak Numfor	9969.0	10.00	13.94	68.00	2
6	Paniai	6535.0	4.20	10.47	65.94	1
7	Puncak Jaya	5459.0	3.51	6.59	64.65	1
8	Mimika	11700.0	9.76	11.77	72.06	2
9	Boven Digoel	8211.0	8.32	10.99	59.16	0
10	Mappi	6268.0	6.29	10.53	64.56	0
11	Asmat	5882.0	4.74	8.47	56.88	0
12	Yahukimo	4737.0	4.01	7.59	65.52	1
13	Pegunungan Bintang	5578.0	2.49	5.79	64.08	1
14	Tolikara	4946.0	3.62	8.04	65.30	1
15	Sarmi	6814.0	8.52	11.55	66.00	2
16	Keerom	8918.0	7.83	12.14	66.35	2
17	Waropen	6978.0	8.87	12.77	65.99	2
18	Supiori	5769.0	8.39	12.72	65.53	2
19	Mamberamo Raya	4755.0	5.46	11.30	57.18	0
20	Nduga	4131.0	0.85	2.95	54.82	3

Gambar 13. Menambahkan atribut kluster ke dataset

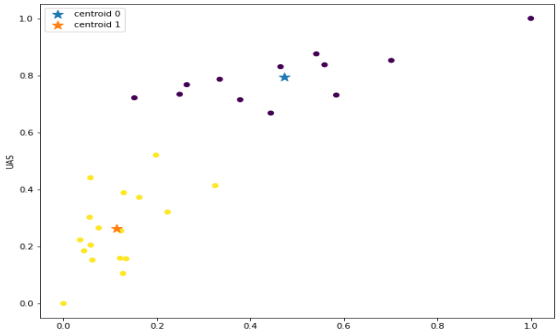
4.4. Evaluasi K Terbaik dengan Metode Elbow

Kita perlu menggunakan metode elbow untuk menemukan nilai cluster terbaik dan optimal. Dengan metode elbow, kita dapat memeriksa apakah nilai pengelompokan (K=4) yang digunakan dalam penelitian ini termasuk dalam klaster terbaik.



Gambar 14. Bagan K Terbaik

Gambar 14 menunjukkan bahwa cluster atau K terbaik dengan metode elbow adalah pada titik persimpangan yang membentuk elbow, yang bernilai 2. Visualisasi data kluster dengan kluster terbaik dapat dilihat pada gambar di bawah ini.



Gambar 15. Visualisasi Pengklusteran Data dengan Klaster Terbaik

Cluster terbaik dengan metode elbow di mana rekomendasi clustering terbaik bernilai 2, sehingga nilai clustering empat diubah menjadi cluster 2. Sebelumnya, pembelajaran tanpa pengawasan adalah dataset yang belum memiliki label. Hasilnya dapat dilihat pada gambar di bawah ini.

	Kabupaten	pengeluaranperkapita	reratalamasekolah	harapanlamasekolah	angkaharapanhidup	cluster
0	Merauke	10430.0	8.49	13.24	66.71	Rendah
1	Jayawijaya	7637.0	5.17	11.58	58.99	Sedang/Baik
2	Jayapura	10160.0	9.60	14.17	66.66	Rendah
3	Nabire	9143.0	9.53	11.14	67.72	Rendah
4	Kepulauan Yapen	7739.0	9.07	12.24	68.85	Sedang/Baik
5	Biak Numfor	9969.0	10.00	13.94	68.00	Rendah
6	Paniai	6535.0	4.20	10.47	65.94	Sedang/Baik
7	Puncak Jaya	5459.0	3.51	6.59	64.65	Sedang/Baik
8	Mimika	11700.0	9.76	11.77	72.06	Rendah
9	Boven Digoel	8211.0	8.32	10.99	59.16	Rendah
10	Mappi	6268.0	6.29	10.53	64.56	Sedang/Baik
11	Asmat	5882.0	4.74	8.47	56.88	Sedang/Baik
12	Yahukimo	4737.0	4.01	7.59	65.52	Sedang/Baik
13	Pegunungan Bintang	5578.0	2.49	5.79	64.08	Sedang/Baik
14	Tolikara	4946.0	3.62	8.04	65.30	Sedang/Baik
15	Sarmi	6814.0	8.52	11.55	66.00	Sedang/Baik
16	Keerom	8918.0	7.83	12.14	66.35	Rendah
17	Waropen	6978.0	8.87	12.77	65.99	Sedang/Baik

Gambar 16. Label baru pada atribut kluster

5. KESIMPULAN

Berdasarkan hasil penelitian, disimpulkan bahwa nilai klaster terbaik yang digunakan dalam penelitian ini adalah 2 dari seluruh klaster 29 kabupaten di Provinsi Papua, di mana 12 kabupaten memiliki IPM rendah. Sebagai perbandingan, 17 kabupaten memiliki IPM sedang (baik). Ke-12 kabupaten tersebut perlu ditingkatkan unsur-unsur penyusun IPM, yaitu pengeluaran per kapita, rata-rata sekolah, harapan sekolah, dan harapan hidup, sehingga dengan bertambahnya unsur-unsur yang membentuk IPM, IPM Provinsi Papua ke depan dapat ditingkatkan. Lebih lanjut, data penelitian ini menunjukkan bahwa clustering dengan K-Means dengan metode elbow menunjukkan bahwa cluster yang terbentuk dapat menghasilkan clustering yang lebih optimal.

DAFTAR PUSTAKA

[1] Y. Pertiwi, D. Permana, N. Amalita, and A. Salma, "Modeling Human Development Index in Papua and West Sumatera with Multivariate Adaptive Regression Spline," *UNP Journal of Statistics and Data Science*, vol. 1, no. 3, pp. 188–195, 2023.

[2] A. Sen, "Issues in the Measurement of Poverty," in *Measurement in public choice*, Springer, 1981, pp. 144–166.

[3] R. F. Ramadhani, H. Yozza, and I. R. HG, "Penerapan Analisis Diskriminan Kuadratik Klasik Untuk Menduga Kategori Indeks Pembangunan Manusia (IPM) Provinsi di Indonesia Tahun 2012," *Jurnal Matematika UNAND*, vol. 4, no. 3, pp. 57–64, 2019.

[4] K. S. Dewi, W. P. Primandhana, and M. Wahed, "Analisis pengaruh pertumbuhan ekonomi, pengangguran, dan kemiskinan terhadap indeks pembangunan manusia di kabupaten bojonegoro," *Syntax Idea*, vol. 3, no. 4, pp. 834–847, 2021.

[5] R. Uspayanti, M. Bawawa, and N. Indriyani, "Students' Interest Towards the Use of Padlet in Grammar Class: A Case in Indonesia," in *SHS Web of Conferences*, 2022, p. 1036.

[6] F. Hardinandar, "Determinan kemiskinan (studi kasus 29 kota/kabupaten di Provinsi Papua)," *Jurnal REP (Riset Ekonomi Pembangunan)*, vol. 4, no. 1, pp. 1–12, 2019.

[7] A. Falakhi, "Pengolahan Data Pelanggan Dengan Tenik Clustering K-Means Di Aplikasi Weka," *Journal Computer Science and Information Systems: J-Cosys*, vol. 3, no. 2, pp. 54–60, 2023.

[8] P. Hariyati, S. Saifullah, and M. Fauzan, "Tehnik Data Mining dalam Mengelompokkan Kasus Pneumonia pada Balita Berdasarkan Provinsi Di Indonesia," *KOMIK (Konferensi Nasional Teknologi Informasi dan Komputer)*, vol. 3, no. 1, 2019.

[9] S. Hendrawan, F. K. S. Dewi, and others, "Clustering Evaluasi Dosen Universitas Atma Jaya Yogyakarta Menggunakan Metode K-Means," *Jurnal Informatika Atma Jogja*, vol. 4, no. 1, pp. 1–8, 2023.

[10] D. P. Kessy and D. R. Romadhoni, "Analisis Hubungan Dinamis Pertumbuhan Ekonomi dan Ketimpangan Pendapatan Indonesia Menggunakan Clustering dan Model Ekonometrik," *Jurnal Matematika dan Statistika serta Aplikasinya Vol*, vol. 9, no. 1, 2021.

[11] R. I. Arumnisaa and A. W. Wijayanto, "Comparison of Ensemble Learning Method: Random Forest, Support Vector Machine, AdaBoost for Classification Human Development Index (HDI)," *SISTEMASI*, vol. 12, no. 1, pp. 206–218, 2023.

[12] Y. N. Dewi, H. Rianto, C. Budihartanti, and F. W. Fibriany, "Penerapan Metode K-Means Dalam Menentukan Kelompok Pendalaman Materi Ujian Nasional," *JISAMAR (Journal of Information System, Applied, Management, Accounting and Research)*, vol. 6, no. 1, pp. 26–31, 2022.

[13] E. H. Bara, Y. A. Pranoto, and F. X. Ariwibisono, "PENGELOMPOKAN DATA OBAT MENGGUNAKAN METODE K-MEANS

CLUSTERING PADA UPT PUSKESMAS KONDORAN KEC. SANGALLA','' *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 4, no. 2, pp. 92–97, 2020.

[14] M. Masruroh and R. Subekti, "Aplikasi regresi partial least square untuk analisis hubungan faktor-faktor yang mempengaruhi indeks pembangunan manusia di Kota Yogyakarta," *Media Statistika*, vol. 9, no. 2, pp. 75–84, 2016.

[15] B. P. Statistik, "Indeks pembangunan manusia," *Retrieved Februari*, vol. 18, 2020.

[16] A. R. Sinaga and G. D. Pranata, "Clustering Data Penjualan Produk pada Toko Yudha dngan Algoritma K-Means," *SisInfo*, vol. 3, no. 2, pp. 135–139, 2021.

[17] A. Solichin and K. Khairunnisa, "Klasterisasi persebaran virus Corona (Covid-19) di DKI Jakarta menggunakan metode K-Means," *Fountain of Informatics Journal*, vol. 5, no. 2, pp. 52–59, 2020.

[18] A. R. Sinaga and G. D. Pranata, "Clustering Data Penjualan Produk pada Toko Yudha dngan Algoritma K-Means," *SisInfo*, vol. 3, no. 2, pp. 135–139, 2021.

[19] H. P. Amalia, Y. Yundari, and H. Helmi, "METODE MAXIMUM LIKELIHOOD DALAM PENAKSIRAN MODEL SPATIAL AUTOREGRESSIVE (Studi Kasus: Indeks Pembangunan Manusia Seluruh Provinsi di Indonesia pada Tahun 2016)," *Bimaster: Buletin Ilmiah Matematika, Statistika dan Terapannya*, vol. 8, no. 3, 2019.