

# Komparasi Algoritma K-Means dan K-Medoids untuk Klasterisasi Data Mining Perjalanan Wisatawan Nusantara Terhadap Jumlah Penduduk

Liza Hesti dan Cut Maisyarah Karyati

Universitas Gunadarma

E-mail: lizabudpar.0902@gmail.com, csyarah@gmail.com

## Abstrak

Indonesia memiliki jumlah penduduk yang sangat besar mencapai lebih dari 278 juta jiwa. Populasi penduduk merupakan modal utama dalam pembangunan kepariwisataan dan menjadi potensi besar dalam perjalanan wisatawan nusantara. Penelitian ini akan melakukan komparasi algoritma K-Means dan K-Medoid dalam pengelompokkan jumlah perjalanan wisatawan nusantara menurut provinsi asal serta korelasinya dengan jumlah penduduk Indonesia. Clustering akan dibagi kedalam 3 (tiga) cluster yaitu tinggi, sedang, rendah. Tool yang digunakan dalam penelitian ini adalah PostgreSQL, Rapidminer, Tableau. Komparasi algoritma dilihat dari performansi vector dan cluster model. Dari hasil komparasi tersebut menunjukkan pengelompokkan 3 provinsi yang memiliki jumlah perjalanan paling tinggi adalah Jawa Tengah, Jawa Barat dan Jawa Timur. Nilai DBI K-Means yakni sebesar 0.381 lebih kecil dari K-Medoid sebesar 1.166 yang berarti bahwa algoritma K-Means lebih baik dari pada algoritma K-Medoids. Hasil clustering divisualisasikan kedalam tableau untuk melihat korelasinya dengan jumlah penduduk. Dashboard tableau menunjukkan adanya korelasi positif antara jumlah penduduk suatu provinsi dengan jumlah perjalanan wisnus.

**Kata kunci:** K-Means, K-Medoids, Wisnus, *Clustering*, *Rapidminer*

## Pendahuluan

Indonesia sebagai negara kepulauan terbesar di dunia dengan lebih dari 17.000 pulau, memiliki jumlah penduduk yang sangat besar, yang menjadikannya sebagai salah satu negara dengan populasi terpadat. Berdasarkan data Badan Pusat Statistik (BPS), pada tahun 2023 jumlah penduduk Indonesia mencapai lebih dari 278 juta jiwa. Sektor pariwisata di Indonesia memiliki peran penting dalam perekonomian negara. Pariwisata tidak hanya berkontribusi terhadap penerimaan devisa negara (PDB) tetapi juga sebagai salah satu upaya dalam pemerataan ekonomi dan pengembangan daerah. Menurut Undang-Undang Republik Indonesia Nomor 10 Tahun 2009[1] tentang Kepariwisata, bahwa kepariwisataan adalah keseluruhan kegiatan yang terkait dengan pariwisata dan bersifat multidimensi serta multidisiplin yang muncul sebagai wujud kebutuhan setiap orang dan negara serta interaksi antara wisatawan dan masyarakat setempat, sesama wisatawan, Pemerintah, Pemerintah Daerah, dan pengusaha.

Jumlah penduduk akan menjadi salah satu modal utama dalam pembangunan kepariwisataan

Wisatawan nusantara atau disingkat dengan wisnus adalah seseorang yang melakukan perjalanan di wilayah teritori suatu negara, dalam hal ini Indonesia, dengan lama perjalanan kurang dari 6 bulan dan bukan bertujuan untuk memperoleh penghasilan ditempat yang dikunjungi serta bukan merupakan perjalanan rutin (sekolah atau bekerja), untuk mengunjungi obyek wisata komersial, dan atau menginap di akomodasi komersial, dan atau jarak perjalanan lebih besar atau sama dengan 100 (seratus) km pergi pulang [2].

Jika dilihat data BPS tahun 2023 terdapat kenaikan jumlah wisatawan nusantara sebesar 12,37% dari tahun sebelumnya[3]. Kenaikan ini ditandai setelah periode pembatasan dan lockdown akibat pandemi COVID-19, banyak masyarakat Indonesia yang mulai merencanakan perjalanan untuk berlibur. Dengan pelonggaran pembatasan dan pemulihan ekonomi, penduduk menjadi lebih berani untuk melakukan perjalanan domestik. Dalam Rencana Strategis (Renstra) Kementerian Pariwisata dan Ekonomi Kreatif/Badan Pariwisata dan Ekonomi Kreatif tahun 2020-2024, pada sasaran strategis ketiga (SS3) [4] yaitu meningkatnya kualitas dan jumlah wisatawan, terdapat salah satu in-

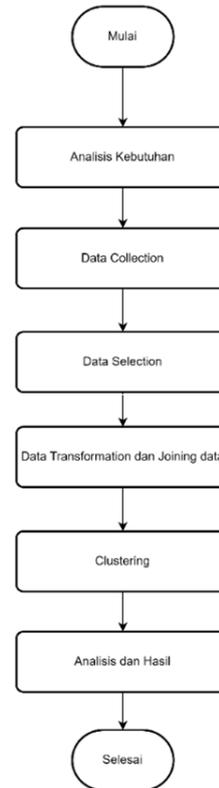
dikator kinerja adalah jumlah wisatawan nusantara.

Publikasi BPS yang berjudul Statistik Wisatawan Nusantara 2023 [3] melihat korelasi antara jumlah perjalanan wisatawan nusantara dengan jumlah penduduk bahwa provinsi dengan jumlah penduduk besar juga cenderung memiliki jumlah perjalanan wisnus yang tinggi. Untuk mengolah dan menganalisis data yang begitu besar, dibutuhkan metode yang efektif. Salah satu metode yang populer adalah klusterisasi. Klusterisasi memungkinkan seseorang untuk mengelompokkan data menjadi beberapa kelompok (cluster) yang homogen, sehingga dapat mengidentifikasi segmen-segmen pasar yang berbeda. Dalam klusterisasi, metode K-Means dan K-Medoids sering digunakan karena kemampuannya untuk mengidentifikasi segmen-segmen pasar yang berbeda dengan cara yang akurat. Hal ini dapat terlihat pada penelitian sebelumnya yang dilakukan oleh Riska dan Farokhah [5] menggunakan data Wisatawan Mancanegara dengan melakukan perbandingan hasil evaluasi menggunakan algoritma K-Means dan K-Medoids untuk mengelompokkan negara berdasarkan jumlah wisatawan mancanegara yang datang ke Indonesia dari yang paling banyak hingga yang paling sedikit. Dan ada pula penelitian dengan menggunakan data Tingkat Kunjungan Wisatawan Mancanegara Berdasarkan Rute Masuk Ke Indonesia dikelompokkan menggunakan algoritma K-Means dan K-Medoids[6]. Algoritma yang sama juga dilakukan perbandingan dalam penelitian yang dilakukan oleh Firtiyadi dan Kurniawati [7] mengelompokkan tingkat kinerja pegawai pada perusahaan perumahan nasional dan membandingkan nilai terbaik antara algoritma K-Means lebih baik dari algoritma K-Medoids.

Menurut Ghazal dkk [8] K-Means merupakan salah satu algoritma yang paling banyak digunakan untuk melakukan clusterisasi. Selain itu Algoritma K-Means juga digunakan oleh Gustian dkk[9] dengan menggunakan data pengunjung pemandian air panas Cikundul di Jawa Barat. Penelitian tersebut mengelompokkan kedalam tiga klaster untuk melihat jumlah pengunjung yang paling banyak sampai sedikit sesuai dengan parameter yang ditetapkan. Hal yang sama juga diteliti oleh Purnomo dkk [10] dalam mengelompokkan objek wisata yang memiliki potensi yang paling rendah dalam kunjungan wisatawan sampai yang paling tinggi. Berdasarkan penelitian terdahulu, algoritma K-Means dan K-Medoids telah banyak digunakan dalam berbagai bidang, termasuk pariwisata. Penelitian ini bertujuan untuk membandingkan performa kedua algoritma tersebut dalam mengelompokkan jumlah perjalanan wisatawan nusantara menurut provinsi asal serta menganalisis hubungan antara jumlah perjalanan wisatawan nusantara dengan jumlah penduduk disetiap provinsi.

## Metode Penelitian

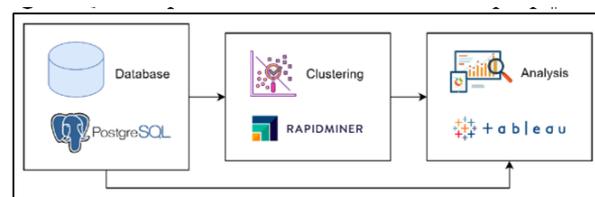
Penelitian ini dilakukan dengan menggunakan algoritma K-Means dan K-Medoids. Adapun tahapan penelitian yaitu analisis kebutuhan, data collection, data selection, data transformation dan joining data, clustering serta analisis dan hasil sesuai Gambar 1



Gambar 1: Tahapan Penelitian

## Analisis Kebutuhan

Pada tahapan ini dilakukan persiapan kebutuhan sistem baik hardware maupun software diantaranya *PostgreSQL*, *Rapidminer*, *Tableau* serta Laptop, lihat Gambar 2. RapidMiner adalah salah satu aplikasi sumber terbuka terdepan dalam penambahan data. Aplikasi ini mengintegrasikan berbagai alat terpisah untuk analisis data dan mesin penambangan data, seperti teknik pengumpulan, transformasi, visualisasi, serta interpretasi data[11].



Gambar 2: Tool Yang Digunakan

### Data Collection

Merupakan tahap proses pengambilan data yang digunakan sebagai referensi dalam penelitian ini. Dalam tahap pengumpulan data, penulis mengambil data sekunder dari BPS yaitu data jumlah perjalanan wisatawan nusantara menurut provinsi asal pertahun mulai dari tahun 2021 sampai dengan

2023, Data jumlah penduduk Indonesia pada tahun 2020-2022 sebanyak 34 provinsi dan data sekunder dari databoks katadata mengenai data jumlah penduduk Indonesia tahun 2023 sebanyak 38 provinsi. Pada website BPS, data jumlah perjalanan wisatawan nusantara menurut provinsi asal disajikan pertahun, lihat Gambar 3 sampai 5.

38 Provinsi	Jumlah Perjalanan Wisatawan Nusantara Menurut Provinsi Asal (Perjalanan)												
	2023												
	Januari	Februari	Maret	April	Mei	Juni	Juli	Agustus	September	Oktober	November	Desember	Tahunan
ACEH	717.217	552.783	598.842	772.003	807.976	713.394	911.767	636.139	588.953	609.838	649.208	746.349	8.304.469
SUMATERA UTARA	2.579.356	1.755.426	1.728.767	2.429.645	2.299.900	2.281.863	2.373.862	1.868.291	1.812.580	1.935.072	1.859.867	2.387.174	25.311.803
SUMATERA BARAT	1.144.057	928.742	968.659	1.078.994	1.179.900	1.172.495	1.227.983	1.010.552	1.009.265	1.042.337	894.678	1.165.678	12.823.340
RIAU	1.049.070	754.324	771.198	1.135.378	1.045.392	1.029.000	1.138.660	932.629	894.379	950.202	816.422	1.041.078	11.557.732
JAMBI	412.204	288.412	296.982	434.081	441.147	436.567	467.650	383.816	368.758	385.164	360.892	429.945	4.705.618
SUMATERA SELATAN	890.809	682.420	681.584	852.145	910.699	943.718	1.058.340	885.872	866.837	904.719	827.240	915.242	10.419.625
BENGKULU	202.639	152.573	157.805	232.020	213.621	214.889	237.589	188.991	184.005	192.386	193.018	210.448	2.379.984
LAMPUNG	1.150.037	943.668	909.180	1.175.047	1.232.696	1.227.576	1.328.185	1.101.053	1.095.330	1.131.906	949.429	1.203.553	13.447.660
KEP. BANGKA BELITUNG	167.336	123.754	134.009	145.833	178.569	185.921	214.356	163.500	172.019	175.878	153.019	177.339	1.991.533
KEP. RIAU	70.945	57.923	62.364	74.387	86.303	91.226	103.446	84.011	79.166	89.935	86.009	82.727	968.442
DKI JAKARTA	6.465.804	5.357.049	4.830.646	4.811.026	5.422.419	5.388.725	6.091.009	5.259.967	5.628.724	5.899.372	5.624.599	5.758.959	66.538.299
JAWA BARAT	16.522.686	13.167.011	11.996.430	13.178.281	13.743.502	12.981.417	13.921.224	11.182.351	13.291.326	13.384.262	13.128.601	14.415.847	160.912.938
JAWA TENGAH	9.692.272	8.645.987	10.184.466	10.093.458	11.216.448	10.557.383	10.625.633	8.184.368	7.957.822	8.503.468	8.208.325	10.488.589	114.358.219
DI YOGYAKARTA	1.937.593	1.736.766	1.955.154	2.097.904	2.151.808	2.049.867	1.816.080	1.474.832	1.488.370	1.478.941	1.422.793	2.046.393	21.656.501
JAWA TIMUR	19.540.342	17.496.798	17.812.321	19.184.554	21.868.127	20.010.407	16.825.288	13.126.883	12.177.131	13.389.819	12.793.788	20.472.978	204.698.436
BANTEN	5.492.473	4.616.258	3.981.837	4.192.130	4.404.233	4.174.715	4.571.762	3.774.769	4.878.498	4.721.935	4.617.568	4.690.474	54.116.652
BALI	1.387.981	937.683	920.118	1.310.802	1.383.827	1.926.214	1.996.939	1.818.231	1.713.019	1.764.857	1.687.893	1.565.915	18.413.479
NUSA TENGGARA BARAT	955.706	662.090	576.779	692.129	954.845	1.299.569	1.343.586	1.106.119	1.042.258	1.117.222	1.107.464	1.070.117	11.927.884
NUSA TENGGARA TIMUR	339.214	204.936	204.908	246.124	263.734	364.548	374.996	306.699	295.820	314.948	305.066	322.823	3.543.816
KALIMANTAN BARAT	406.156	294.234	287.401	357.941	396.288	407.204	405.514	307.553	283.267	312.082	335.912	403.279	4.196.831
KALIMANTAN TENGAH	272.356	207.831	212.955	232.575	279.838	292.933	310.831	263.410	239.006	253.894	211.471	281.744	3.058.844
KALIMANTAN SELATAN	672.905	535.300	536.224	607.919	672.932	697.547	766.405	628.972	601.944	623.230	528.647	681.245	7.553.270

Gambar 3: Data Wisnus Menurut Provinsi Asal Thn 2023

**Jumlah Penduduk Menurut Provinsi di Indonesia (Ribu Jiwa), 2020-2022**

[← back](#)
[📄 xlsx](#)

Data series subyek **Kependudukan** juga dapat diakses melalui **Fitur Tabel Dinamis**.

Data Series : 2020-2022 2018-2019

Search:

Provinsi di Indonesia	Jumlah Penduduk Menurut Provinsi di Indonesia (Ribu Jiwa)		
	2020	2021	2022
Aceh	5 274,9	5 333,7	5 407,9
Sumatera Utara	14 799,4	14 936,2	15 115,2
Sumatera Barat	5 534,5	5 580,2	5 640,6
Riau	6 394,1	6 493,6	6 614,4
Kep. Riau	2 064,6	2 118,2	2 179,8
Jambi	3 548,2	3 585,1	3 631,1
Sumatera Selatan	8 467,4	8 550,9	8 657,0
Kep. Bangka Belitung	1 455,7	1 473,2	1 494,6
Bengkulu	2 010,7	2 032,9	2 060,1
Lampung	9 007,8	9 081,8	9 176,6
DKI Jakarta	10 562,1	10 609,7	10 680,0

Gambar 4: Data Jumlah Penduduk 2020-2022 (34 Provinsi)



Gambar 5: Data Jumlah Penduduk Tahun 2023 (38 Provinsi)

### Data Selection

Data selection berarti memilih data yang relevan dari dataset yang lebih besar, fokus pada atribut atau variabel yang berkaitan langsung dengan tujuan penelitian. Data yang didapatkan pada proses pengumpulan data, selanjutnya diseleksi dan dikelola didalam Database. Tabel yang dibuat pada tahapan ini adalah:

1. Tabel perjalanan\_wisatawan\_2021, terdiri atas 34 rows.
2. Tabel perjalanan\_wisatawan\_2022, terdiri atas 34 rows
3. Tabel perjalanan\_wisatawan\_2023, terdiri atas 34 rows
4. Tabel jumlah\_penduduk\_2020\_2022, terdiri atas 34 rows
5. Tabel jumlah\_penduduk\_2023, terdiri atas 38 rows

### Data Transformation dan Joining Data

Data Transformation adalah proses mengubah data ke dalam format yang sesuai untuk analisis. Tahap ini bisa melibatkan normalisasi data, pembuatan variabel baru, atau agregasi data. Joining Data adalah menggabungkan data dari beberapa tabel.

Tabel 1: Tabel Joining Data Wisnus 2021-2023

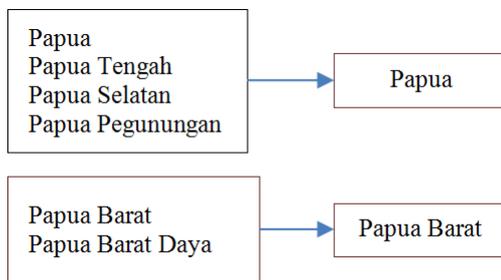
perjalanan.perjalanan\_wisatawan/postgres/postgres@PostgreS...

Query: `SELECT * FROM perjalanan.perjalanan_wisatawan`

Provinsi	2021	2022	2023	
text	numeric	numeric	numeric	
1	ACEH	5632315	7483957	8304469
2	BALI	9009667	11132389	18413479
3	BANTEN	38396859	48935825	54116652
4	BENGKULU	1639658	1988998	2379984
5	DI YOGYAKARTA	24551526	18888517	21656501
6	DKI JAKARTA	51705258	63081040	66538299
7	GORONTALO	1299319	1619987	1960412
8	JAMBI	3066614	3892005	4705618
9	JAWA BARAT	97358488	128667116	160912938
10	JAWA TENGAH	134782286	103991668	114358219
11	JAWA TIMUR	158616085	198913339	204698436
12	KALIMANTAN BARAT	2271124	3138799	4196831
13	KALIMANTAN SELATAN	4506915	5716978	7553270
14	KALIMANTAN TENGAH	1727916	2438971	3058844
15	KALIMANTAN TIMUR	2365615	5266467	7099201
16	KALIMANTAN UTARA	229797	334625	498643
17	KEP. BANGKA BELITUNG	1005293	1586000	1991533
18	KEP. RIAU	504707	782185	968442
19	LAMPUNG	8855256	10745577	13447660
Total rows: 34 of 34		Query complete 00:00:00.271		

Pada mulanya data jumlah perjalanan wisatawan nusantara yang bersumber dari BPS berbentuk data perprovinsi per tahun, sedangkan untuk kebutuhan analisis, data akan di transformasi dalam bentuk series 2021-2023 per provinsi. Oleh karena itu akan dilakukan proses joining data yaitu data jumlah perjalanan wisatawan nusantara menurut provinsi asal tahun 2021, 2022 dan tahun 2023, lihat Tabel 1.

Dikarenakan data jumlah penduduk tahun 2023 terdapat 38 provinsi, sedangkan data jumlah perjalanan wisatawan nusantara berdasarkan Data BPS yang diakses pada bulan April-Mei terdiri dari 34 provinsi maka dilakukan transformasi data yakni data 38 provinsi menjadi 34 provinsi.



Gambar 6: Transformasi Data

Pada Gambar 6, Provinsi Papua, Papua Tengah, Papua Selatan, Papua Pegunungan dilakukan transformasi menjadi Provinsi Papua. Hal yang sama juga dilakukan transformasi data untuk provinsi Papua Barat dan Papua Barat Daya menjadi provinsi Papua Barat.

Salah satu Teknik dalam transformasi adalah menggunakan normalisasi min-max dengan rumus (1) :

$$v' = \frac{v - \min A}{\max A - \min A} (new_{\max A} - new_{\min A}) + new_{\min A} \quad (1)$$

$v'$  = Nilai hasil normalisasi

$v$  = Nilai yang dinormalkan

$\min A$  = Nilai terendah (minmal)

$\max A'$  = Nilai tertinggi (maksimal)

$new_{\min A}$  = Nilai minimal baru, 0

$new_{\max A}$  = Nilai maksimal baru, 1

Hasil normalisasi Tabel 2 dengan Min-Max disimpan di dalam Database sesuai Tabel 3.

Tabel 2: Jumlah Penduduk 2021-2023

	Provinsi text	2021 numeric	2022 numeric	2023 numeric
1	ACEH	5333700	5407900	5515839
2	BALI	4362700	4415100	4344554
3	BANTEN	12061500	12252000	12469997
4	BENGKULU	2032900	2060100	2098089
5	DI YOGYAKARTA	3712900	3761900	3722296
6	DKI JAKARTA	10609700	10680000	11337563
7	GORONTALO	1181000	1192700	1237185
8	JAMBI	3585100	3631100	3760275
9	JAWA BARAT	48782400	49405800	49899992
10	JAWA TENGAH	36742500	37032400	38125191
11	JAWA TIMUR	40878800	41150000	41644099
12	KALIMANTAN BARAT	5470800	5541400	5557277
13	KALIMANTAN SELATAN	4122600	4182100	4234214
14	KALIMANTAN TENGAH	2702200	2741100	2753049
15	KALIMANTAN TIMUR	3808200	3859800	4007736
16	KALIMANTAN UTARA	713600	727800	747415
17	KEP. BANGKA BELITUNG	1473200	1494600	1521723
18	KEP. RIAU	2118200	2179800	2178610

Tabel 3: Normalisasi Jumlah Perjalanan Wisnus 2021-2023

	Provinsi text	2021_Norm numeric	2022_Norm numeric	2023_Norm numeric
1	ACEH	0.0341	0.0360	0.0382
2	BALI	0.0554	0.0544	0.0877
3	BANTEN	0.2410	0.2447	0.2626
4	BENGKULU	0.0089	0.0083	0.0092
5	DI YOGYAKARTA	0.1536	0.0934	0.1036
6	DKI JAKARTA	0.3250	0.3160	0.3234
7	GORONTALO	0.0068	0.0065	0.0072
8	JAMBI	0.0179	0.0179	0.0206
9	JAWA BARAT	0.6132	0.6463	0.7856
10	JAWA TENGAH	0.8495	0.5220	0.5576
11	JAWA TIMUR	1.0000	1.0000	1.0000
12	KALIMANTAN BARAT	0.0129	0.0141	0.0181
13	KALIMANTAN SELATAN	0.0270	0.0271	0.0345
14	KALIMANTAN TENGAH	0.0095	0.0106	0.0125
15	KALIMANTAN TIMUR	0.0135	0.0248	0.0323
16	KALIMANTAN UTARA	0.0000	0.0000	0.0000
17	KEP. BANGKA BELITUNG	0.0049	0.0063	0.0073
18	KEP. RIAU	0.0017	0.0023	0.0023
19	LAMPUNG	0.0545	0.0524	0.0634

Pada tahapan ini telah dilakukan:

1. penggabungan data jumlah perjalanan wisatawan nusantara tahun 2021-2023 per provinsi

dalam bentuk agregat.

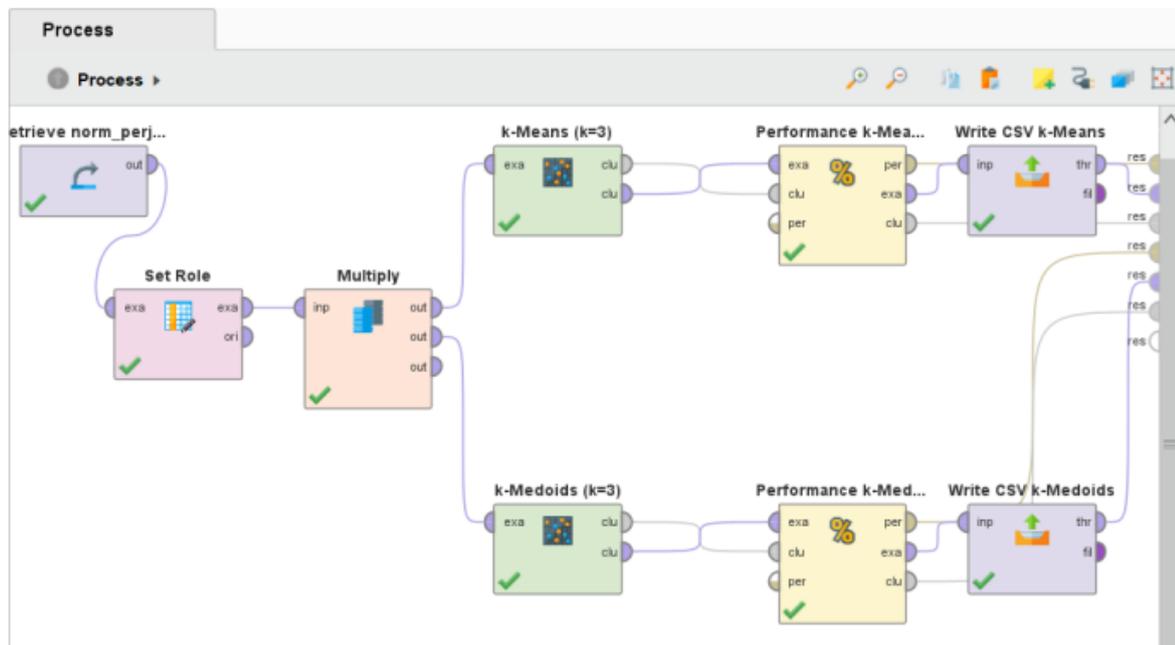
2. penggabungan provinsi pada data jumlah penduduk Indonesia yang sebelumnya dari 38 provinsi menjadi 34 provinsi mengikuti jumlah provinsi pada data jumlah perjalanan wisatawan nusantara menurut provinsi asal
3. normalisasi data pada tabel jumlah perjalanan wisatawan nusantara tahun 2021-2023.

### Clustering

Pada tahap ini dilakukan proses clustering (lihat Gambar 7) data dengan menggunakan algoritma

K-Means dan K-Medoids untuk data jumlah perjalanan wisatawan menurut provinsi asal. Tahapan pengujian dalam clustering yakni:

1. siapkan dataset yang akan digunakan.
2. Set role
3. pilih parameter K-Means dan K-Medoid.
4. set k=3 dan max run=10, dan max optimization steps = 100
5. lakukan validasi kinerja masing-masing algoritma menggunakan performance vector dengan konsep *Davies-Bouldin Index*



Gambar 7: Proses Clustering

Tabel yang digunakan saat proses clustering adalah tabel normalisasi perjalanan wisnus 2021-2023. Dilakukan koneksi data dengan Database khususnya tabel normalisasi perjalanan wisatawan. Operator yang digunakan dalam pengambilan data adalah retrieve data. Kemudian dilakukan pemilihan operator multiply karena algoritma yang akan digunakan lebih dari satu yaitu K-Means dan K-Medoids. Operator selanjutnya adalah operator K-Means dan K-Medoids dengan parameter k=3 dikarenakan akan dibagi kedalam 3 kategori yakni tinggi, sedang dan rendah serta menggunakan parameter performance vector yaitu Davies Bouldien Index. Hasil Clustering di simpan dalam bentuk csv menggunakan operator wite csv.

Hasil Clustering dalam bentuk csv disimpan dan dikelola dalam Database. Tabel yang telah terbentuk diolah di database membentuk tabel baru yang akan digunakan untuk analisis berikutnya, lihat Tabel 4 sampai 8.

Tabel 4: centroid\_K-Means

	Provinsi [PK] text	cluster text	2021_Centroid numeric	2022_Centroid numeric	2023_Centroid numeric
1	ACEH	cluster_0	0.027	0.032	0.036
2	BALI	cluster_0	0.027	0.032	0.036
3	BANTEN	cluster_2	0.283	0.280	0.293
4	BENGKULU	cluster_0	0.027	0.032	0.036
5	DI YOGYAKARTA	cluster_0	0.027	0.032	0.036
6	DKI JAKARTA	cluster_2	0.283	0.280	0.293
7	GORONTALO	cluster_0	0.027	0.032	0.036
8	JAMBI	cluster_0	0.027	0.032	0.036
9	JAWA BARAT	cluster_1	0.821	0.723	0.781
10	JAWA TENGAH	cluster_1	0.821	0.723	0.781
11	JAWA TIMUR	cluster_1	0.821	0.723	0.781
12	KALIMANTAN BARAT	cluster_0	0.027	0.032	0.036
13	KALIMANTAN SELATAN	cluster_0	0.027	0.032	0.036
14	KALIMANTAN TENGAH	cluster_0	0.027	0.032	0.036
15	KALIMANTAN TIMUR	cluster_0	0.027	0.032	0.036
16	KALIMANTAN UTARA	cluster_0	0.027	0.032	0.036
17	KEP BANGKA BELITUNG	cluster_0	0.027	0.032	0.036
18	KEP RIAU	cluster_0	0.027	0.032	0.036
19	LAMPUNG	cluster_0	0.027	0.032	0.036

Total rows: 34 of 34 Query complete 00:00:00.164

Tabel 5: Centroid\_K-Medoids

	Provinsi [PK] text	cluster text	2021_Centroid numeric	2022_Centroid numeric	2023_Centroid numeric
1	ACEH	cluster_2	0.040	0.039	0.049
2	BALI	cluster_2	0.040	0.039	0.049
3	BANTEN	cluster_0	0.105	0.109	0.121
4	BENGKULU	cluster_2	0.040	0.039	0.049
5	DI YOGYAKARTA	cluster_0	0.105	0.109	0.121
6	DKI JAKARTA	cluster_0	0.105	0.109	0.121
7	GORONTALO	cluster_2	0.040	0.039	0.049
8	JAMBI	cluster_2	0.040	0.039	0.049
9	JAWA BARAT	cluster_1	1	1	1
10	JAWA TENGAH	cluster_1	1	1	1
11	JAWA TIMUR	cluster_1	1	1	1
12	KALIMANTAN BARAT	cluster_2	0.040	0.039	0.049
13	KALIMANTAN SELATAN	cluster_2	0.040	0.039	0.049
14	KALIMANTAN TENGAH	cluster_2	0.040	0.039	0.049
15	KALIMANTAN TIMUR	cluster_2	0.040	0.039	0.049
16	KALIMANTAN UTARA	cluster_2	0.040	0.039	0.049
17	KEP BANGKA BELITUNG	cluster_2	0.040	0.039	0.049
18	KEP. RIAU	cluster_2	0.040	0.039	0.049
19	LAMPUNG	cluster_2	0.040	0.039	0.049

Tabel 8: Data\_gabungan\_scatterplot

	Provinsi text	avg_penduduk numeric	avg_perjalanan numeric	k-Means text	k-Medoids text
1	ACEH	5419146	7140247	cluster_0	cluster_2
2	BALI	4374118	12851845	cluster_0	cluster_2
3	BANTEN	12261166	47149779	cluster_2	cluster_0
4	BENGKULU	2063696	2002880	cluster_0	cluster_2
5	DI YOGYAKARTA	3732365	21698848	cluster_0	cluster_0
6	DKI JAKARTA	10875754	60441532	cluster_2	cluster_0
7	GORONTALO	1203628	1626573	cluster_0	cluster_2
8	JAMBI	3658825	3888079	cluster_0	cluster_2
9	JAWA BARAT	49362731	128979514	cluster_1	cluster_1
10	JAWA TENGAH	37300030	117710724	cluster_1	cluster_1
11	JAWA TIMUR	41224300	187409287	cluster_1	cluster_1
12	KALIMANTAN BARAT	5523159	3202251	cluster_0	cluster_2
13	KALIMANTAN SELATAN	4179638	5925721	cluster_0	cluster_2
14	KALIMANTAN TENGAH	2732116	2408577	cluster_0	cluster_2
15	KALIMANTAN TIMUR	3891912	4910428	cluster_0	cluster_2
16	KALIMANTAN UTARA	729605	354355	cluster_0	cluster_2
17	KEP BANGKA BELITUNG	1496508	1527609	cluster_0	cluster_2
18	KEP. RIAU	2158870	751778	cluster_0	cluster_2
19	LAMPUNG	9103286	11016164	cluster_0	cluster_2

Total rows: 34 of 34 Query complete 00:00:00.244

Tabel 6: Visualisasi\_K-Means

	Provinsi text	Tahun Integer	Jumlah Penduduk numeric	Cluster text	Centroid numeric
1	ACEH	2021	5333700	cluster_0	0.027
2	ACEH	2022	5407900	cluster_0	0.032
3	ACEH	2023	5515839	cluster_0	0.036
4	BALI	2021	4362700	cluster_0	0.027
5	BALI	2022	4415100	cluster_0	0.032
6	BALI	2023	4344554	cluster_0	0.036
7	BANTEN	2021	12061500	cluster_2	0.283
8	BANTEN	2022	12252000	cluster_2	0.280
9	BANTEN	2023	12469997	cluster_2	0.293
10	BENGKULU	2021	2032900	cluster_0	0.027
11	BENGKULU	2022	2060100	cluster_0	0.032
12	BENGKULU	2023	2098089	cluster_0	0.036
13	DI YOGYAKARTA	2021	3712900	cluster_0	0.027
14	DI YOGYAKARTA	2022	3761900	cluster_0	0.032
15	DI YOGYAKARTA	2023	3722296	cluster_0	0.036
16	DKI JAKARTA	2021	10609700	cluster_2	0.283
17	DKI JAKARTA	2022	10680000	cluster_2	0.280
18	DKI JAKARTA	2023	11337563	cluster_2	0.293
19	GORONTALO	2021	1181000	cluster_0	0.027

Total rows: 102 of 102 Query complete 00:00:00.189

Tabel 7: Visualisasi\_K-Medoids

	Provinsi text	Tahun Integer	Jumlah Penduduk numeric	Cluster text	Centroid numeric
1	ACEH	2021	5333700	cluster_2	0.040
2	ACEH	2022	5407900	cluster_2	0.039
3	ACEH	2023	5515839	cluster_2	0.049
4	BALI	2021	4362700	cluster_2	0.040
5	BALI	2022	4415100	cluster_2	0.039
6	BALI	2023	4344554	cluster_2	0.049
7	BANTEN	2021	12061500	cluster_0	0.105
8	BANTEN	2022	12252000	cluster_0	0.109
9	BANTEN	2023	12469997	cluster_0	0.121
10	BENGKULU	2021	2032900	cluster_2	0.040
11	BENGKULU	2022	2060100	cluster_2	0.039
12	BENGKULU	2023	2098089	cluster_2	0.049
13	DI YOGYAKARTA	2021	3712900	cluster_0	0.105
14	DI YOGYAKARTA	2022	3761900	cluster_0	0.109
15	DI YOGYAKARTA	2023	3722296	cluster_0	0.121
16	DKI JAKARTA	2021	10609700	cluster_0	0.105
17	DKI JAKARTA	2022	10680000	cluster_0	0.109
18	DKI JAKARTA	2023	11337563	cluster_0	0.121
19	GORONTALO	2021	1181000	cluster_2	0.040

Total rows: 102 of 102 Query complete 00:00:00.134

## Analisis dan Hasil

Pada tahapan ini akan dilakukan analisis komparasi dan hasil clustering menggunakan algoritma K-Means dan K-Medoids.

## Perbandingan Performance Vector

PerformanceVector (K-Means)	PerformanceVector (K-Medoids)
PerformanceVector: Avg. within centroid distance: 0.012 Avg. within centroid distance_cluster_0: 0.004 Avg. within centroid distance_cluster_1: 0.009 Avg. within centroid distance_cluster_2: 0.004 Davies Bouldin: 0.381	PerformanceVector: Avg. within centroid distance: 0.030 Avg. within centroid distance_cluster_0: 0.039 Avg. within centroid distance_cluster_1: 0.256 Avg. within centroid distance_cluster_2: 0.002 Davies Bouldin: 1.166

Gambar 8: Hasil Performance Vector K-Means dan K-Medoid

Jika dilihat dari performance vector yang dihasilkan menunjukkan nilai Davies Bouldin menggunakan algoritma K-Means lebih kecil dari Davies Bouldin K-Medoids. Nilai Davies-Bouldin Index (DBI) yang lebih rendah menunjukkan performa klasterisasi yang lebih baik karena mengindikasikan bahwa cluster lebih kompak (lebih dekat satu sama lain dalam clusternya) dan lebih terpisah dari cluster lainnya. K-Means memiliki nilai DBI yang lebih rendah (0.381), yang berarti performa clustering K-Means lebih baik dibandingkan dengan K-Medoids (1.166). Oleh karena itu, K-Means merupakan algoritma dengan performa clustering yang lebih baik dalam penelitian ini.

## Perbandingan Cluster Model

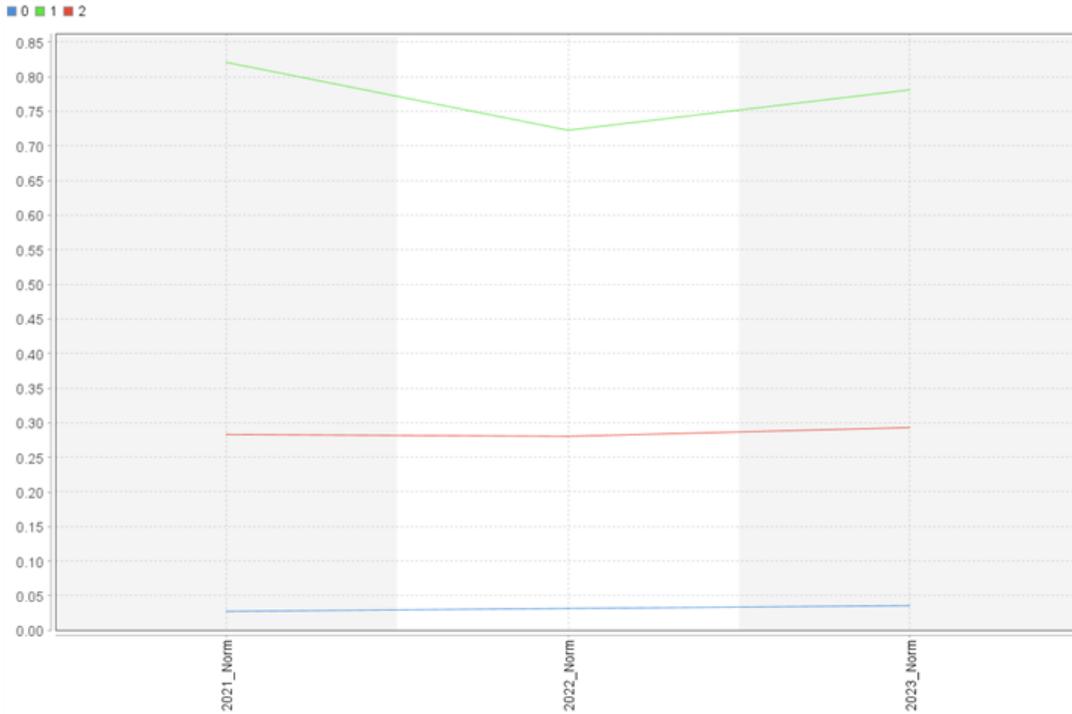
Hasil cluster model dapat dilihat pada Gambar 9 sampai 11 dan Tabel 9 dan Tabel 10.

Cluster Model (K-Means)	Cluster Model (K-Medoids)
Cluster 0: 29 items	Cluster 0: 5 items
Cluster 1: 3 items	Cluster 1: 3 items
Cluster 2: 2 items	Cluster 2: 26 items
Total number of items: 34	Total number of items: 34

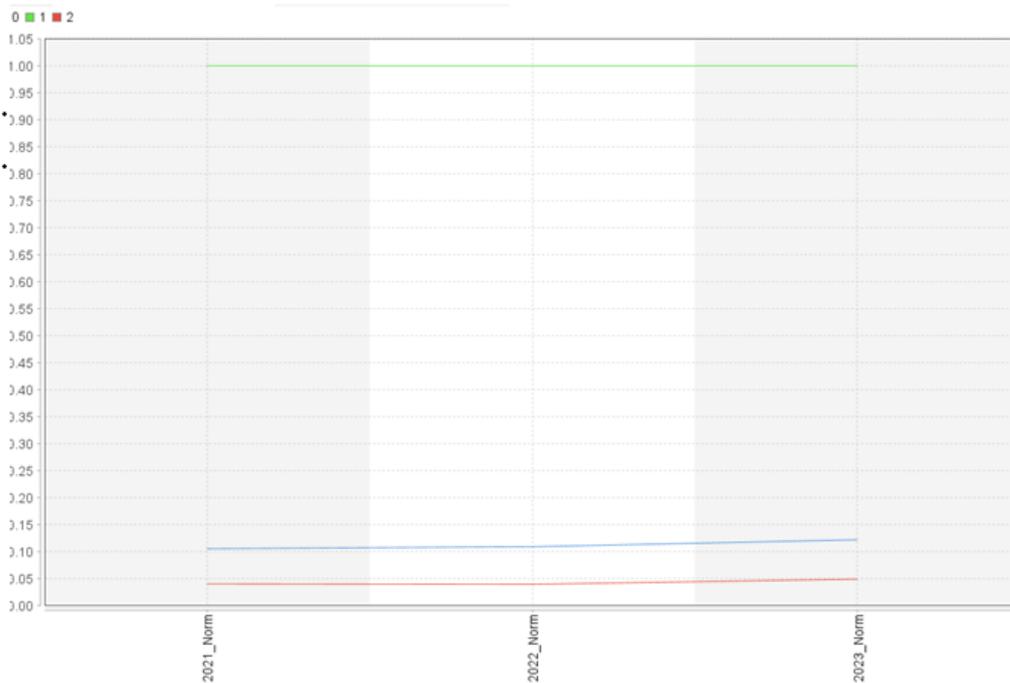
Tabel 9: Centroid K-Means

Attribute	cluster_0	cluster_1	cluster_2
2021_Norm	0.027	0.821	0.283
2022_Norm	0.032	0.723	0.280
2023_Norm	0.036	0.781	0.293

Gambar 9: Cluster Model K-Means dan K-Medoids



Gambar 10: Plot K-Means



Gambar 11: Plot K-Medoids

Tabel 10: Centroid K-Medoids

Attribute	cluster_0	cluster_1	cluster_2
2021_Norm	0.105	1	0.040
2022_Norm	0.109	1	0.039
2023_Norm	0.121	1	0.049

Centroid K-Means pada Tabel 9 diartikan:

1. Cluster\_0: mewakili kelompok wisatawan dengan jumlah perjalanan yang sangat sedikit atau bahkan hampir tidak ada pada ketiga tahun tersebut. Dibuktikan dengan nilai-nilai centroid yang sangat kecil.
2. Cluster\_1: merupakan kelompok terbesar dengan jumlah perjalanan yang paling signifikan pada setiap tahun. Dibuktikan dengan nilai centroid yang jauh lebih besar dibandingkan dengan cluster lainnya.
3. Cluster\_2: Kelompok ini kelompok sedang dari jumlah perjalanan wisatawan. Jumlah perjalanannya lebih sedikit dibandingkan cluster\_1, tetapi lebih banyak dibandingkan cluster\_0.

Hasil centroid Tabel 9 dapat dilihat pada visualisasi plot sesuai Gambar 10. Warna hijau menunjukkan cluster\_1, warna merah menunjukkan cluster\_2, dan warna biru menunjukkan cluster\_0.

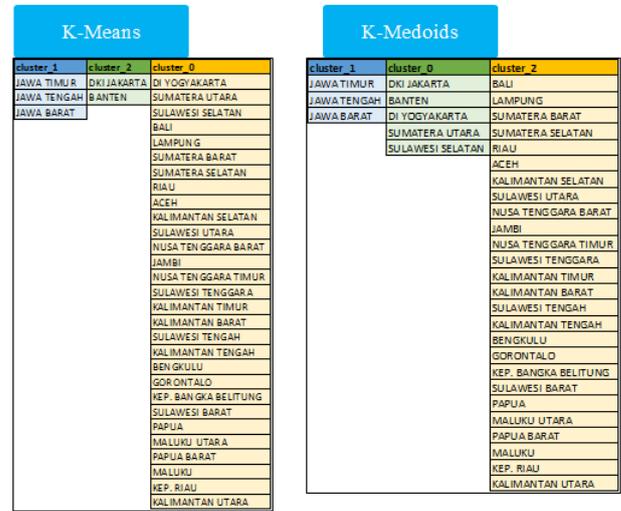
Centroid K-Medoids pada tabel 10 diartikan:

1. Cluster\_2: mewakili kelompok wisatawan dengan jumlah perjalanan yang sangat sedikit atau bahkan hampir tidak ada pada ketiga tahun tersebut. Dibuktikan dengan nilai-nilai centroid yang sangat kecil.
2. Cluster\_1: merupakan kelompok terbesar dengan jumlah perjalanan yang paling signifikan pada setiap tahun. Dibuktikan dengan nilai centroid yang jauh lebih besar dibandingkan dengan cluster lainnya.
3. Cluster\_0: Kelompok ini kelompok sedang dari jumlah perjalanan wisatawan. Jumlah perjalanannya lebih sedikit dibandingkan cluster\_1, tetapi lebih banyak dibandingkan cluster\_2.

Hasil centroid Tabel 10 dapat dilihat pada visualisasi plot sesuai Gambar 11. Warna hijau menunjukkan cluster\_1, warna biru menunjukkan cluster\_0, dan warna merah menunjukkan cluster\_2.

Dari hasil perbandingan cluster model diatas, terlihat bahwa algoritma K-Means terdiri atas 3 cluster, cluster 1 adalah kategori tinggi sebanyak 3 provinsi, cluster 2 adalah kategori sedang sebanyak 2 provinsi dan cluster 0 adalah kategori rendah sebanyak 29 provinsi. Sedangkan algoritma K-Medoids menghasilkan 3 cluster, dengan cluster 1 merupakan cluster tinggi yang terdiri atas

3 provinsi, cluster 0 merupakan cluster sedang sebanyak 5 provinsi dan cluster 2 merupakan cluster rendah sebanyak 26 provinsi. Perbandingan clustering berdasarkan provinsi dapat dilihat pada Gambar 12.



Gambar 12: Perbandingan Clustering per Kategori

Setelah melakukan analisis clustering perbandingan algoritma K-Means dan K-Medoids baik dari sisi performance vector maupun cluster model dari jumlah perjalanan wisatawan nusantara menurut provinsi asal, akan dilakukan juga analisis korelasi jumlah wisatawan nusantara menurut provinsi asal dengan jumlah penduduk.

1. Hubungan Titik Centroid dari Hasil Clustering Terhadap Perubahan Jumlah Penduduk Sepanjang Tahun 2021 – 2023

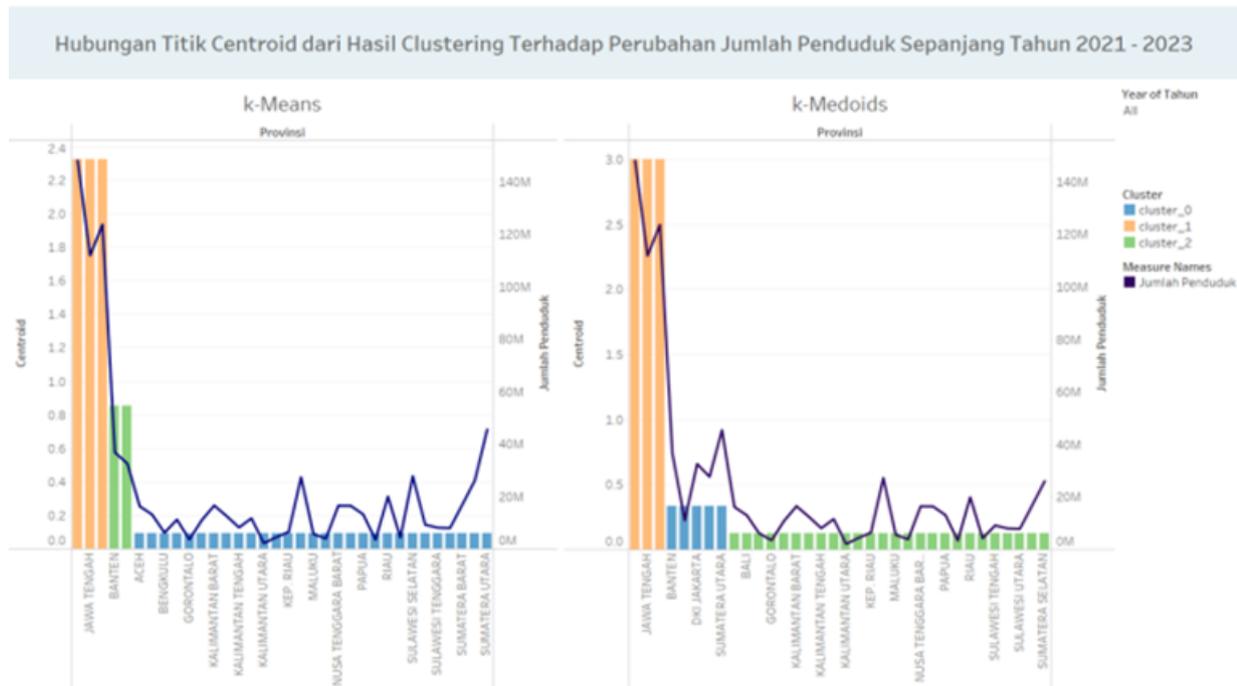
Gambar 13 menunjukkan perbedaan hasil clustering berdasarkan dua metode berbeda dan bagaimana centroid dari kluster-kluster tersebut berhubungan dengan jumlah penduduk di setiap provinsi. Jawa Tengah, Jawa Barat, dan Jawa Timur di kedua metode clustering berada di cluster 1 yang memiliki nilai centroid yang lebih tinggi, menunjukkan bahwa kedua provinsi ini memiliki karakteristik serupa dalam hal jumlah penduduk yang cukup tinggi dibandingkan provinsi lainnya.

2. Korelasi antara Nilai Rata-rata Jumlah Penduduk dan Jumlah Perjalanan berdasarkan Hasil Clustering Tahun 2021 – 2023

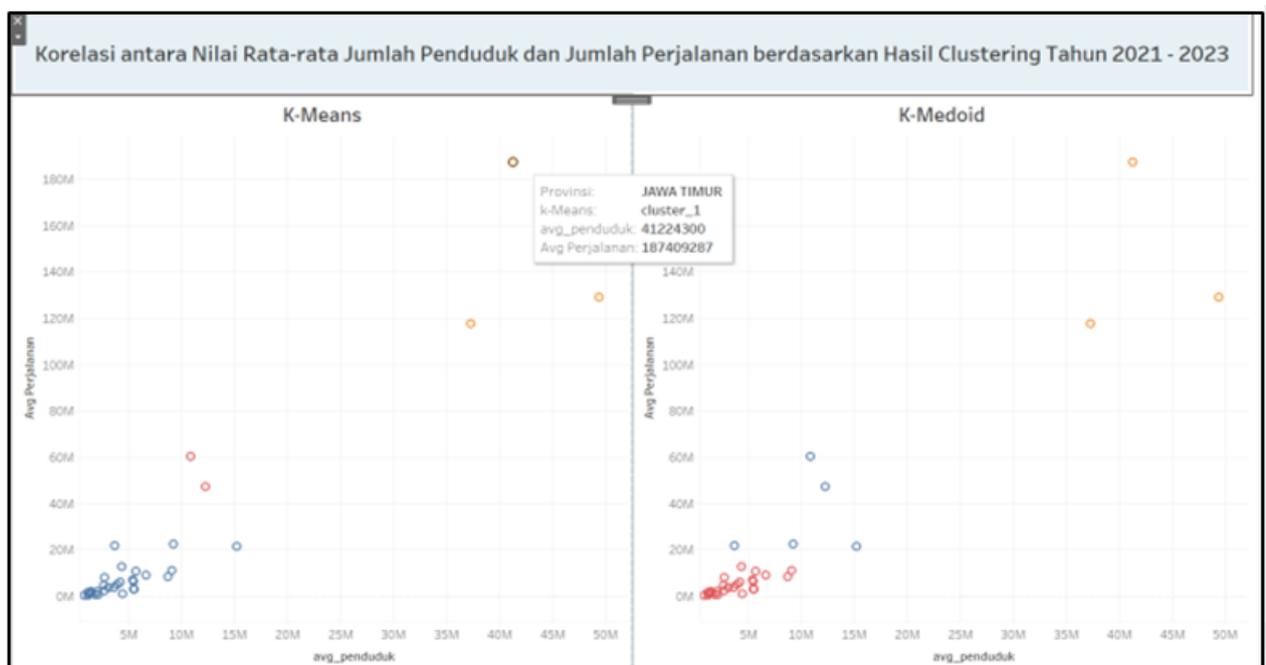
Gambar 14 menampilkan scatterplot yang menggambarkan korelasi antara nilai rata-rata jumlah penduduk dan jumlah perjalanan wisatawan nusantara menurut provinsi asal berdasarkan hasil clustering untuk tahun 2021-2023. Scatterplot ini dibagi menjadi dua bagian berdasarkan dua metode clustering, yaitu K-Means (panel kiri) dan K-Medoids (panel kanan). Panel K-Means (kiri): a) titik orange menunjukkan provinsi dengan jumlah penduduk rata-rata yang lebih tinggi dan jumlah perjalanan wisata yang lebih tinggi. Provinsi-provinsi ini cenderung memiliki angka perjalanan

wisata yang lebih sedikit. b) titik merah menunjukkan provinsi dengan jumlah penduduk rata-rata sedang dan jumlah perjalanan wisata kelompok

sedang. c) titik biru menunjukkan provinsi dengan jumlah penduduk rata-rata rendah dan jumlah perjalanan wisata kelompok rendah.



Gambar 13: Grafik Hubungan Centroid-Jumlah Penduduk



Gambar 14: Scatterplot Korelasi Rata-rata Jumlah Penduduk dengan Jumlah Perjalanan Wisnus

Panel K-Medoids (kanan): a) titik orange menunjukkan provinsi dengan jumlah penduduk rata-rata yang lebih tinggi dan jumlah perjalanan wisata yang lebih tinggi. Provinsi-provinsi ini cen-

derung memiliki angka perjalanan wisata yang lebih sedikit. b) titik biru menunjukkan provinsi dengan jumlah penduduk rata-rata sedang dan jumlah perjalanan wisata kelompok sedang. c) titik merah me-

nunjukkan provinsi dengan jumlah penduduk rata-rata rendah dan jumlah perjalanan wisata kelompok rendah.

Kedua metode clustering ini menunjukkan korelasi positif antara rata-rata jumlah penduduk dan jumlah perjalanan wisata, yang artinya semakin tinggi jumlah penduduk suatu provinsi, semakin besar juga jumlah perjalanan wisata yang tercatat. Provinsi dengan populasi penduduk yang lebih besar cenderung memiliki jumlah perjalanan wisatatawan nusantara yang lebih tinggi, terlihat dari beberapa titik oranye yang berada jauh di sisi kanan atas grafik. Secara keseluruhan, grafik ini membantu menunjukkan bagaimana metode K-Means dan K-Medoids mengelompokkan provinsi berdasarkan data penduduk dan perjalanan wisatawan, serta menyoroti hubungan antara kedua variabel tersebut.

## Penutup

Berdasarkan hasil penelitian ini dapat disimpulkan bahwa algoritma K-Means lebih unggul dibandingkan K-Medoids dalam pengelompokan data perjalanan wisatawan nusantara menurut provinsi asal tahun 2021-2023, yang ditunjukkan melalui nilai Davies Bouldin Index (DBI). Pada K-Means, hasil klusterisasi mengidentifikasi 3 provinsi dengan tingkat perjalanan wisata tertinggi (Jawa Tengah, Jawa Barat, dan Jawa Timur), 2 dua provinsi dengan tingkat sedang, dan 29 provinsi dengan tingkat rendah. Sebaliknya, K-Medoids menghasilkan 3 provinsi pada kluster tertinggi, 5 provinsi pada tingkat sedang, dan 26 provinsi pada tingkat rendah. Selain itu, analisis klusterisasi ini menunjukkan adanya korelasi positif antara jumlah penduduk suatu provinsi dengan jumlah perjalanan wisatawan nusantara, yang tampak jelas pada visualisasi scatterplot, di mana provinsi dengan populasi lebih besar cenderung memiliki jumlah perjalanan yang lebih tinggi. Hasil klusterisasi pada penelitian ini memberikan rekomendasi kepada pemerintah dan pemangku kepentingan pariwisata agar provinsi dengan tingkat perjalanan wisata sedang dan rendah mendapat perhatian lebih, dengan merancang strategi promosi yang intensif mengenai destinasi wisata di Indonesia melalui media digital, iklan televisi, dan media sosial untuk meningkatkan kesadaran masyarakat terhadap potensi wisata di sekitar mereka. Provinsi dengan tingkat perjalanan sedang dan rendah perlu mendapat perhatian untuk merancang strategi promosi yang lebih intensif guna menarik wisatawan termasuk dengan memanfaatkan media digital dan kampanye wisata yang lebih efektif. Adapun saran untuk peneliti selanjutnya adalah mengelompokkan data jumlah perjalanan wisatawan nusantara menurut provinsi tujuan, sehingga dapat melihat provinsi yang paling banyak dikunjungi dan dikorelasikan dengan objek wisata yang ada di provinsi tersebut

serta menggunakan tool Orange, WEKA atau KNIME dan menggunakan algoritma clustering selain K-Means dan K-Medoids.

## Daftar Pustaka

- [1] Anonim, Undang-undang (UU) Nomor 10 Tahun 2009 tentang Kepariwisata”, Republik Indonesia, 2009.
- [2] N. Yuniati, “Profil dan Karakteristik Wisatawan Nusantara (studi kasus di Yogyakarta)”, *J. Pariwisata Pesona*, vol. 3, no. 2, pp. 175–190, doi: 10.26905/jpp.v3i2.2381, 2018.
- [3] Anonim, “Statistik Wisatawan Nusantara 2023”, Nomor Katalog : 8401009, Badan Pusat Statistik, Mei 2024.
- [4] Anonim, “Rencana Strategis Kemenparekraf/Baparekraf 2020-2024”, Kementerian Pariwisata dan Ekonomi/ Badan Pariwisata dan Ekonomi Kreatif, Agustus 2020.
- [5] S. Y. Riska dan L. Farokhah, “Perbandingan Hasil Evaluasi Algoritma K-Means dan K-Medoid Berdasarkan Kunjungan Wisatawan Mancanegara ke Indonesia”, *INTEGER J. Inf. Technol.*, vol. 8, pp. 1–8, 2023.
- [6] R. A. Nugraha, Y. Cahyana, dan A. R. Juwita, “Penerapan Algoritma K-Means dan K-Medoid untuk Mengelompokkan Tingkat Kunjungan Wisatawan Asing Berdasarkan Jalur Masuk ke Indonesia”, *Sci. Student J. Information, Technol. Sci.*, vol. V, no. 1, pp. 112–122, 2024.
- [7] A. U. Fitriyadi dan A. Kurniawati, “Analisis Algoritma K-Means dan K-Medoids Untuk Clustering Data Kinerja Karyawan Pada Perusahaan Perumahan Nasional”, *KILAT*, vol. 10, no. 1, pp. 157–168, doi: 10.33322/kilat.v10i1.1174, 2021.
- [8] Taher M. Ghazal, Muhammad Zahid Hussain, Raed A. Said, Afrozah Nadeem, Mohammad Kamrul Hasan, Munir Ahmad, Muhammad Adnan Khan, and Muhammad Tahir Naseem, “Performances of K-Means Clustering Algorithm with Different Distance Metrics”, *Intell. Autom. Soft Comput.*, vol. 29, no. 3, pp. 735–742, doi: 10.32604/iasc.2021.019067, 2021.
- [9] D. Gustian, M. Muslih, D. A. Dewi, V. Siti, R. Rustiana, and Purnamawati, “K-Means In Clustering The Satisfaction Level Of Cikundul Hot Water”, *J. Theor. Appl. Inf. Technol.*, vol. 101, no. 17, pp. 6980–6990, 2023.
- [10] B. Setio Purnomo dan P. Taqwa Prasetyaningrum, “Penerapan Data Mining Dalam Mengelompokkan Kunjungan Wisatawan Di Kota

Yogyakarta Menggunakan Metode K-Means”,  
J. Comput. Sci. Technol., vol. 1, no. 1, pp.  
27–32, doi: 10.54840/jcstech.v1i1.38 Sep. 2022.

Data Mining Dalam Mengelompokkan Kun-  
jungan Wisatawan Mancanegara Di Prov. Su-  
lawesi Selatan Dengan K-Means Dan SVM”,  
J. Inform. J. Pengemb. IT, vol. 8, no. 3, pp.  
174–180, doi: 10.30591/jpit.v8i3.4554, 2023.

[11] N. C. Gosari dan R. Rismayani, “Penerapan