Perbandingan Hasil Evaluasi Algoritma K-Means dan K-Medoid Berdasarkan Kunjungan Wisatawan Mancanegara ke Indonesia

Suastika Yulia Riska[1], Lia Farokhah[2]
[1]Prodi Teknik Informatika, Fakultas Teknologi dan Desain, Institut Teknologi dan Bisnis Asia Malang
[2]Prodi Teknik Informatika, Fakultas Teknologi dan Desain, Institut Teknologi dan Bisnis Asia Malang
Email: [1]riska.suastika@asia.ac.id, [2]farokhah@asia.ac.id

Abstract*. Tourism is one of the activities that can affect the level of the economy of a country. Tourism activities that can attract foreign tourists to enter Indonesia are one of the income contributions for the State of Indonesia. Based on data from BPS, it shows a significant decrease in the number of foreign tourists coming to Indonesia. The purpose of this study is to form a group of countries based on the number of foreign tourists entering Indonesia from the most to the fewest. Then the results of the grouping can be used as a reference for the government to implement the right strategy to increase the interest of tourists coming to Indonesia. The algorithm used for the clustering process is K-Means and K-Medoids, by applying the values of k=2, k=3, and k=5. The evaluation process used the Davies Bouldin index method. By using the K-Medoids Algorithm, the best cluster results are using k = 2 with a davies bouldin index -0.567. The best cluster result in this case is to use the K-Measn Algorithm with k=5, and with the result the Davies bouldin index value of -0.302.*
Keywords: *Data Mining, Clustering, K-Means, K-Medoid, International Travelers*

Abstrak. *Kegiatan Pariwisata merupakan salah satu kegiatan yang dapat mempengaruhi tingkat perekonomian suatu negara. Kegiatan pariwisata yang dapat menarik minat wisatawan mancanegara masuk ke Indonesia menjadi salah satu kontribusi pendapatan untuk Negara Indonesia. Berdasarkan data dari BPS menunjukkan adanya penurunan yang signifikan dari jumlah wisatawan mancanegara yang datang ke Indonesia. Tujuan penelitian ini adalah untuk membentuk kelompok negara-negara berdasarkan jumlah wisatawan mancanegara yang masuk ke Indonesia dari yang paling banyak hingga yang paling sedikit. Kemudian hasil pengelompokkan tersebut dapat digunakan sebagai acuan pemerintah untuk menerapkan strategi yang tepat untuk meningkatkan antusias wisatawan datang ke Indonesia. Algoritma yang digunakan untuk proses clustering adalah K-Means dan K-Medoids, dengan menerapkan nilai k=2, k=3, dan k=5. Proses evaluasi digunakan metode davies bouldin index. Dengan menggunakan Algoritma K-Medoids, hasil cluster terbaik adalah dengan menggunakan k=2 dengan nilai davies bouldin index -0,567. Hasil cluster terbaik dalam kasus ini adalah dengan menggunakan Algoritma K-Means dengan k=5, dan dengan hasil nilai davies bouldin index -0.302.*
Kata Kunci: *Data Mining, Clustering, K-Means, K-Medoid, Wisatawan Mancanegara*

1. Pendahuluan

Pariwisata merupakan salah satu aspek yang dapat memberikan kontribusi pendapatan pada suatu daerah. Adanya pendapatan daerah dan masyarakat dari kegiatan pariwisata dapat meningkatkan kesejahteraan masyarakat, karena masyarakat mendapatkan kesempatan untuk membuka dan memperluas usaha. Pembangunan pariwisata merupakan salah satu pembangunan yang perlu dikembangkan karena dari sektor ini dapat meningkatkan penerimaan devisa negara, menghasilkan pertumbuhan ekonomi yang cepat dalam menyediakan lapangan kerja, peningkatan penghasilan, standar hdup serta menstimulasikan faktor-faktor produksi yang lainnya (Urbanus & Febianti, 2017).

Wisatawan yang dibedakan menjadi 2, yaitu wisatawan lokal dan wisatawan mancanegara. Wisatawan lokal merupakan masyarakat yang melakukan kegiatan pariwisata tanpa keluar dari batas negara sendiri, sedangkan wisatawan mancanegara merupakan masyarakat yang melakukan kegiatan pariwisata dengan cara mendatangi negara lain. Dengan adanya wisatawan mancanegara yang datang ke Indonesia, maka dapat meningkatkan perekonomian negara Indonesia. Selain itu, negara Indonesia dapat memperkenalkan keindahan pariwisata dan *culture* negara Indonesia.

ISSN: 2579-566X (Online)
ISSN: 2477-5274 (Print)

Berdasarkan data yang diperoleh dari Badan Pusat Statistik (BPS), jumlah kunjungan wisatawan mancanegara ke Indonesia selama tahun 2021 mengalami penurunan sebesar 61,57%, yaitu sebesar 1,56 juta kunjungan dibandingkan selama tahun 2020. Hal tersebut merupakan penurunan yang sangat signifikan dibandingkan dengan sebelum adanya pancemi Covid-19. Kunjungan wisatawan mancanegara ke Indonesia selama Tahun 2019 sebanyak 16,1 juta wisatawan, dan pada Tahun 2020 terjadi penurunan yang sangat signifikan, yaitu 4 juta wisatawan (Jannah, 2022).

Adanya penurunan tersebut perlu adanya *cluster*isasi negara-negara berdasarkan jumlah wisatawan yang datang ke Indonesia. Sehingga, dapat diketahui warga dari negara mana saja yang sangat kurang antusias atau sangat tinggi antusiasnya untuk datang ke Indonesia sehingga pemerintah Indonesia bisa membuat strategi yang tepat untuk peningkatan antusias warga mancanegara datang ke Indonesia dengan cara menganalisis data kluster yang terbentuk. Dengan menerapkan Algoritma-Algoritma dalam data mining, dapat dihasilkan *clustering* tersebut. Pada penelitian ini diterapkan 2 Algoritma *clustering*, yaitu K-Means dan K-Medoid. Tujuannya adalah untuk membandingkan hasil evaluasi kedua algoritma tersebut untuk mengelompokkan negara berdasarkan jumlah wisatawan mancanegara yang datang ke Indonesia. Kedua Algoritma ini sudah banyak digunakan untuk *clustering* dalam berbagai kasus.

Pada penelitian yang dilakukan oleh (Irfiani et al., 2018) , tentang *clustering* untuk menentukan Gizi Balita menggunakan Algoritma K-Means. *Clustering* dilakukan menjadi 5 kategori, yaitu obesitas, gizi lebih, gizi baik, gizi kurang, dan gizi buruk. Adanya kategori tersebut dapat digunakan sebagai penanganan dini pihak posyandu dalam menanggulangi gizi buruk dan obesitas. Sehingga, dapat ditemukan terdapat 30% balita obesitas dan 11% kekurangan gizi. Pada penelitian yang dilakukan oleh (Alfatah, 2021) tentang *clustering* pemetaan strategi daerah pemilih pada calon legislative DPR RI menggunakan Algoritma K-Means dengan membentuk 3 kategori yaitu daerah tingkat pemilih sedang , tingkat pemilih tinggi, dan tingkat pemilih rendah. Penelitian yang dilakukan oleh (Farokhah & Aditya, 2017) tentang *clustering* rekomendasi tema tugas akhir mahasiswa menggunakan Algoritma K-Means menjadi 7 kategori. Hasilnya menunjukkan hasil centroid kedua memiliki nilai kebenaran terbaik dengan nilai 90,24%. Penelitian yang dilakukan oleh (Mustofa & Iman Saufik Suasana, 2020) tentang *clustering* menggunakan K-Medoids pada e-Government bidang Information and Communication Technology dalam penentuan Status EDGI *(E- Government Development Index)*. *Clustering* dilakukan dengan membuat 4 kategori. Adapun K-Medoid mampu menghasilkan nilai Bouldin Index yang optimal yaitu 0,593. Penelitian yang dilakukan oleh (Riyanto, 2019) tentang *clustering* penyebaran diare di Kota Medan menggunakan Algoritma K-Medoid dengan membentuk 2 kategori yaitu daerah yang masyarakatnya sering terkena diare, dan daerah yang masyarakatnya jarang terkena diare. Sehingga, data tersebut dapat digunakan oleh pemerintah sebagai acuan dalam memberikan perhatian kepada kluster daerah yang masyarakatnya sering terkena diare. Penelitian yang dilakukan oleh (Alodia et al., 2021) tentang *clustering* provinsi di Indonesia berdasarkan indikator pendidikan menggunakan K-Medoids menjadi 3 kategori, yaitu indikator pendidikan tinggi, indikator pendidikan sedang, dan indikator pendidikan rendah. Penelitian yang dilakukan oleh (Dwilestari et al., 2021) tentang *clustering* data penduduk miskin di Indonesia menggunakan Algoritma K-Medoid. Dari hasil pengelompokan dengan 3 kategori, diperoleh *cluster* presentase kemiskinan tingkat rendah sebanyak 19 provinsi, *cluster* presentase kemiskinan tingkat sedang sebanyak 12 provinsi dan *cluster* presentase kemiskinan tingkat tinggi sebanyak 3 provinsi.

Pada penelitian ini dilakukan perbandingan evaluasi penggunaan Algoritma K-Means dan K-Medoid untuk menentukan *cluster* negara berdasarkan jumlah kunjungan wisatawan yang datang ke Indonesia. Bagi *cluster* negara yang tergolong rendah, dapat digunakan oleh Pemerintah sebagai acuan untuk meningkatkan promosi agar jumlah wisatawan negara tersebut untuk berkunjung ke Indonesia dapat meningkat. Selain itu, bagi negara dengan jumlah wisatawan terbanyak bisa dibuat acuan pemerintah dengan membuat kebijakan yang makin memudahkan bagi negara tersebut sehingga semakin meningkatkan minat. Pada akhirnya, pemerintah bisa meningkatkan potensi pariwisata di Indonesia.

2. Tinjauan Pustaka

2.1 Data Mining

Data mining didefinisikan sebagai proses menemukan pola-pola pada data dalam jumlah besar menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat. Adapun tahapan-tahapan data mining meliputi :

1. Pembersihan Data, yaitu proses menghilangkan noise dan data yang tidak konsisten atau data tidak relevan.
2. Integrasi Data, yaitu penggabungan data dari berbagai database ke dalam satu database baru.
3. Seleksi Data, yaitu data yang ada pada database sering kali tidak semuanya dipakai, oleh karena itu hanya data yang sesuai untuk dianalisis yang akan diambil dari database.
4. Transformasi Data, yaitu data diubah atau digabung ke dalam format yang sesuai untuk diproses dalam Data Mining.
5. Proses Mining, yaitu suatu proses utama saat metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data.
6. Evaluasi Pola, yaitu untuk mengidentifikasi pola-pola menarik kedalam knowledge based yang ditemukan. (Alfatah, 2021).

Data mining sangat diperlukan dalam mengelola data yang sangat besar. Tujuannya adalah untuk memudahkan aktivitas recording dalam suatu transaksi dan untuk proses data worehousing agar dapat memberikan informasi yang akurat bagi penggunanya. Data mining sangat banyak diminati dalam industri informasi, karena adanya data dalam jumlah yang sangat besar. Sehingga data tersebut dapat diolah menjadi informasi yang penting dan dapat dimanfaatkan oleh masyarakat (Riyanto, 2019).

2.2 *Clustering*

Clustering merupakan proses pengelompokan data dalam beberapa kategori yang mengakibatkan data dalam satu kategori memiliki kemiripan yang tinggi. *Clustering* tidak diberikan label, hal tersebut mengakibatkan seluruh atribut yang dimiliki dianggap sama (Mustofa & Iman Saufik Suasana, 2020). Metode *clustering* banyak dikembangkan oleh para ahli, sehingga memiliki karakter, kelebihan, dan kekurangan masing-masing. Menurut kategori kekompakan, pengelompokan terbagi menjadi dua, yaitu komplet dan parsial. Data dinamakan data yang kompak menjadi satu kategori jika semua data dapat dikelompokkan menjadi satu. Namun jika terdapat sedikit data yang tidak masuk dalam satu kelompok disebut dengan data noise (Dinata et al., 2020).

2.3 K-Means

Algoritma *K-Means* merupakan salah satu metode *clustering* dengan masukan data tanpa label kelas (unsupervised learning) yang digunakan untuk mengelompokkan data ke dalam dua atau lebih kelompok. Kelompok Data yang memiliki karakteristik sama, dikelompokkan dalam satu *cluster*, dan data yang memiliki karakteristik berbeda dikelompokan dengan *cluster* yang berbeda. Sehingga data yang berada dalam satu *cluster* memiliki tingkat variasi yang kecil (Alfatah, 2021). Langkah Algoritma K-Means antara lain (Irfiani et al., 2018):

1. Tentukan k sebagai jumlah *cluster* yang di bentuk. Penentuan banyaknya jumlah *cluster* k dilakukan dengan beberapa faktor seperti pertimbangan teoritis dan konseptual yang diusulkan untuk menentukan berapa banyak *cluster*.
2. Bangkitkan k Centroid (titik pusat *cluster*) awal secara random. Untuk menentukan centroid awal dilakukan secara acak dari beberapa objek yang tersedia sebanyak k *cluster*, untuk menghitung centroid *cluster* ke-i berikutnya, menggunakan rumus sebagai berikut:

$$v = \frac{\sum_{i=1}^{n} x_i}{n} \quad ; i = 1,2,3,\ldots . n \qquad (1)$$

   dimana : v = centroid pada *cluster*; Xi = objek ke-I; n = banyaknya objek atau jumlah objek yang menjadi anggota *cluster*
3. Hitung jarak setiap objek ke masing-masing centroid dari masing-masing *cluster*. Kemudian hitung jarak antara objek dengan centroid, dalam penelitian ini menggunakan Euclidian Distance

$$d(x, y) = \|x - y\| = \sqrt{\sum (xi - yi)^2} \;\; i = 1,2,3, \ldots n$$

   dimana : xi = objek x ke-i; y = daya y ke-i; n = banyaknya objek.

ISSN: 2579-566X (Online)
ISSN: 2477-5274 (Print)

4. Alokasikan masing-masing objek ke dalam centroid yang paling terdekat.
5. Lakukan iterasi, kemudian tentukan posisi centroid baru dengan menggunakan persamaan
6. Ulangi langkah 3 jika posisi centroid baru tidak sama

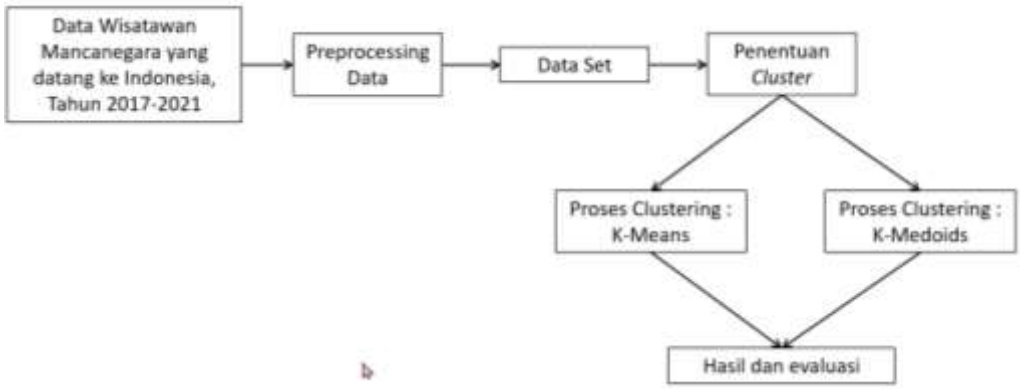
2.4 K-Medoids

Algoritma K-Medoids cukup efisien untuk dataset yang kecil. Langkah awal K-Medoid adalah mencari titik yang paling representatif (medoids) dalam sebuah dataset dengan menghitung jarak dalam kelompok dari semua kemungkinan kombinasi dari medoids sehingga jarak antar titik dalam suatu *cluster* kecil sedangkan jarak titik antar *cluster* besar (Dwilestari et al., 2021). Algoritma K-Medoids memiliki kelebihan untuk mengatasi kelemahan pada pada algoritma K-Means yang sensitive terhadap noise dan outlier, dimana objek dengan nilai yang besar yang memungkinkan menyimpang pada dari distribusi data. Kelebihan lainnya yaitu hasil proses *clustering* tidak bergantung pada urutan masuk dataset. Langkah- langkah algoritma K-Medoids:

1. Inisialisasi pusat *cluster* sebanyak k (jumlah *cluster*)
2. Alokasikan setiap data (objek) ke *cluster*terdekat menggunakan persamaan ukuran jarak Euclidian Distance dengan persamaan: $(x, y) = \|x - y\| = \sqrt{\sum(xi - yi)^2} \;\; ni = 1,2,3, \ldots n$
3. Pilih secara acak objek pada masing-masing *cluster* sebagai kandidat medoid baru
4. Hitung jarak setiap objek yang berada pada masing-masing *cluster* dengan kandidat medoidbaru.
5. Hitung total simpangan (S)dengan menghitungnilai total distancebaru –total distancelama. Jika S< 0, maka tukar objek dengan data *cluster* untuk membentuk sekumpulan kobjek baru sebagai medoid.
6. Ulangi langkah 3 sampai 5 hingga tidak terjadi perubahan medoid, sehingga didapatkan *cluster* beserta anggota *cluster* masing-masing. (Riyanto, 2019)

3. Metode Penelitian

Data yang digunakan pada penelitian ini diperoleh dari BPS tentang data kunjungan wisatawan mancanegara yang datang ke Indonesia pada tahun 2017-2021. Adapun data negara yang digunakan dibatasi pada wilayah negara-negara yang termasuk dalam Benua Asia. Algoritma yang digunakan pada penelitian ini adalah Algoritma K-Means dan K-Medoid, dengan membandingkan kedua tersebut. Selain itu pada penelitian ini juga melakukan *clustering* untuk membentuk beberapa *cluster*, yaitu 2, 3, dan 5 *cluster*. Kerangka penelitian ditunjukkan pada Gambar 1.



Gambar 1. Kerangka Penelitian

Berdasarkan Gambar 1 dimulai dengan tahap preprocessing data. Adapun data yang digunakan adalah nama negara Benua Asia dan jumlah pengunjung setiap tahun, yaitu pada tahun 2017-2021. Adapun daftar Negara yang digunakan dalam proses *clustering* ditunjukkan pada Tabel 1.

Tabel 1. Daftar Negara Benua Asia

No.	*Nama Negara*	*No.*	*Nama Negara*	*No.*	*Nama Negara*	*No.*	*Nama Negara*
1	Brunei Darussalam	13	South Korea	25	Mongolia	37	Uni Emirat Arab

No.	*Nama Negara*	*No.*	*Nama Negara*	*No.*	*Nama Negara*	*No.*	*Nama Negara*
2	Malaysia	14	Pakistan	26	Maladewa	38	Yaman
3	Philippines	15	Bangladesh	27	Nepal	39	Qatar
4	Singapore	16	Srilanka	28	Tajikhistan	40	Irak
5	Thailand	17	Taiwan	29	Timor Leste	41	Iran
6	Vietnam	18	China	30	Turki	42	Israel
7	Laos	19	Afganistan	31	Turkmenistan	43	Jordan
8	Kamboja	20	Bhutan	32	Uzbekistan	44	Libanon
9	Myanmar/Burma	21	Kazakhstan	33	Saudi Arabia	45	Oman
10	Hong Kong	22	Kirgistan	34	Bahrain	46	Palestina
11	India	23	Korea Utara	35	Kuwait	47	Syiria
12	Japan	24	Macao	36	Mesir (Egypt)		

Pada tahap *preprocessing* dilakukan pemilihan data negara yang termasuk ke dalam negara pada Benua Asia. Sehingga, diperoleh data yang dibutuhkan pada tahap *clustering*. Pada proses penentuan *cluster* penelitian ini melakukan analisis dengan membentuk 3 uji coba, yaitu dengan menerapkan K=2, K=3, dan K=5. Ketiga uji coba tersebut diterapkan pada masing-masing Algoritma, yaitu K-Means dan K-Medoids. Pada tahap hasil dan evaluasi ditentukan nilai K yang terbaik untuk proses *clustering* dengan menggunakan Metode *davies bouldin index (DBI)*. Setelah diketahui nilai *DBI* setiap *cluster*, maka nilai DBI yang terkecil menunjukkan hasil *cluster* yang paling baik dan menunjukkan jumlah *cluster* yang optimal (Muningsih et al., 2021).

Dengan mengetahui kelompok negara yang sering ataupun jarang datang ke Indonesia, dapat digunakan sebagai acuan Pemerintah untuk membuat strategi dalam meningkatkan antusias wisatawan mancanegara untuk datang ke Indonesia.

4. Hasil dan Pembahasan

Pada penelitian ini menggunakan perbandingan dua algoritma *clustering* yaitu K-Means dan K-Medoids. K-Means dan K-Medoid merupakan Algoritma pengklasteran berbasis jarak yang membagi data ke dalam sejumlah *cluster*. Perbedaan keduanya ada pada pemilihan titik pusat *cluster*. Nilai evaluasi *davies bouldin index* terhadap jumlah kluster menggunakan algoritma K-Means ditunjukkan pada Tabel 2.

Tabel 2. Nilai evaluasi davies bouldin index terhadap jumlah kluster menggunakan algoritma K-Means

Jumlah Cluster	*Evaluasi Davies Bouldin Index*
2	-0,328
3	-0,756
5	-0.302

Dari hasil tabel 2 maka, jumlah kluster terbaik adalah data dibagi menjadi 5 kelompok kluster. Evaluasi menggunakan *davies bouldin index* memiliki rujukan semakin kecil nilainya semakin baik kelompok kluster tersebut. Berdasarkan pengelompokan menjadi 2 *cluster* menggunakan Algoritma K-Means diketahui bahwa jumlah wisatawan dari *cluster* 1 (*cluster*_0) sebanyak 43 negara dan *cluster* 2 (*cluster*_1) sebanyak 4 negara. Negara-negara yang masuk dalam *cluster* 2 antara lain : Malaysia, Singapore, China, dan Timor Leste. Indek *cluster* pada rapidminer dimulai dari angka nol. Kemudian, hasil analisis titik centroid kedua kluster ditunjukkan pada Tabel 3.

Tabel 3. Titik Centroid K-Means dengan k=2

Attribute	*Cluster_0*	*Cluster_1*
2021	1281	343407
2020	12025	623742
2019	65942	2041414
2018	62536	2043422
2017	68793	1682301

Berdasarkan Tabel 3, diketahui bahwa *cluster*_1 merupakan kluster dengan negeri kunjungan terbanyak. Disisi lain, pada *cluster*_0 memiliki jumlah kunjungan negara luar negeri yang lebih sedikit.

ISSN: 2579-566X (Online)
ISSN: 2477-5274 (Print)

Hasil Analisa Ketika kunjungan wisatawan dibagi menjadi 3 kluster memiliki nilai evaluasi paling kecil yaitu -0,756 dimana itu artinya tidak disarankan mengkluster kunjungan wisatawan mancanegara menjadi 3 kluster karena akan menghasilkan informasi yang salah. Berdasarkan hasil *clustering* k=3, diketahui bahwa jumlah wisatawan yang dikluster menjadi 3 terdiri dari *cluster_0* (*cluster* 1) sebanyak 43 negara, *cluster_1* sebanyak 2 negara, dan *cluster_2* sebanyak 2 negara. Negara yang termasuk pada *cluster*_1 adalah Malaysia dan China, sedangkan Negara yang masuk dalam kelompok *cluster_2* adalah Singapore dan Timor Leste. Setelah itu, hasil analisis titik centroid kedua kluster ditunjukkan pada Tabel 4.

Tabel 4. Titik Centroid K-Means dengan k=3

Attribute	*Cluster_0*	*Cluster_1*	*Cluster_2*
2021	1281	267718	419096
2020	12025	609943	637541
2019	65942	2526416	1556413
2018	62536	2321252	1765593
2017	68793	2107529	1257072

Berdasarkan Tabel 4, diketahui bahwa *cluster_1* merupakan kelompok negara dengan wisatawan terbanyak datang ke Indonesia, diikuti dengan *cluster_2* dan *cluster*_0 sebagai jumlah kunjungan paling sedikit. Pada *clustering* data kunjungan wisatawan menjadi 5 kelompok. Berdasarkan hasil *clustering* k=5, diketahui bahwa jumlah wisatawan yang dikelompokkan menjadi 5 *cluster* terdiri dari *cluster_0* sebanyak 38 negara. *Cluster_1* sebanyak 2 negara yaitu Singapore dan China. *Cluster_2* sebanyak 1 negara yaitu Malaysia. *Cluster_3* sebanyak 1 negara yaitu Timor Leste. *Cluster_4* sebanyak 5 negara yaitu Philippines, India, Japan, South Korea, dan Taiwan. Indek *cluster* pada rapidminer dimulai dari angka nol. Pembagian k=5 merupakan *cluster* terbaik. Pada Tabel 5 menunjukkan titik centroid masing-masing *cluster*.

Tabel 5. Titik Centroid K-Means dengan k=5

Attribute	*Cluster_0*	*Cluster_1*	*Cluster_2*	*Cluster_3*	*Cluster_4*
2021	584	36708	480723	819488	6578
2020	3986	260130	980118	994590	73121
2019	21100	2003262	2980753	1178381	406741
2018	20467	1953952	2503344	1762442	382257
2017	22407	1823645	2121888	960026	421331

Berdasarkan Tabel 5 diketahui bahwa *cluster* dengan kunjungan terbanyak berturut-turut adalah *cluster*_3, *cluster*_2, *cluster*_1, *cluster*_4 dan *cluster*_0. Nilai evaluasi *davies bouldin index* terhadap jumlah *cluster* menggunakan algoritma K-Medoids ditunjukkan pada Tabel 6.

Tabel 6 Nilai evaluasi davies bouldin index terhadap jumlah kluster menggunakan algoritma K-Medoids

Jumlah Kluster	*Evaluasi Davies Bouldin Index*
2	-0,567
3	-0,986
5	-0.916

Berdasarkan Tabel 6, maka jumlah *cluster* terbaik adalah data dibagi menjadi 2 *cluster*. Evaluasi menggunakan *Davies Bouldin Index* memiliki arti bahwa semakin kecil nilainya semakin baik kelompok kluster tersebut. Berdasarkan proses *clustering* dengan 2 *cluster*, maka diketahui bahwa jumlah wistaawan dengan *cluster_0* sebanyak 4 negara dan *cluster_1* sebanyak 43 negara. Adapun Negara yang termasuk dalam *cluster_0* adalah Malaysia, Singapore, China, dan Timor Leste. Hasil analisis titik centroid kedua kluster ditunjukkan pada tabel 7.

Tabel 7. Titik Centroid K-Medoids dengan k=2

Attribute	*Cluster_0*	*Cluster_1*
2021	819488	182

2020	994590	428
2019	1178381	2322
2018	1762442	1941
2017	960026	2328

Berdasarkan Tabel 7, diketahui bahwa *cluster_0* merupakan kluster terbanyak kunjungannya diikuti dengan *cluster_1*.

Pada proses *clustering* data kunjungan wisatawan mancanegara ke Indonesia dikelompokkan menjadi 3 kelompok menggunakan algoritma K-Medoids. Berdasarkan proses *clustering* diketahui bahwa *cluster_1* sebanyak 36 negara. *Cluster_0* sebanyak 4 negara yaitu Malaysia, Singapore, China, dan Timor Leste. *Cluster_2* sebanyak 7 negara yaitu Philippines, Thailand, India, Japan, South Korea, Taiwan, dan Saudi Arabia. Hasil analisis titik centroid ditunjukkan pada Tabel 8. Hasilnya menunjukkan bahwa *cluster_0* merupakan kluster terbanyak kunjungannya diikuti dengan *cluster_2* dan *cluster_1*.

Tabel 8 Titik Centroid K-Medoids dengan k=3

Attribute	*Cluster_0*	*Cluster_1*	*Cluster_2*
2021	819488	182	2053
2020	994590	428	31906
2019	1178381	2322	157512
2018	1762442	1941	165912
2017	960026	2328	182086

Pada proses *clustering* data kunjungan wisatawan menjadi 5 kelompok, menggunakan algoritma k-medoids. Sehingga, diketahui bahwa jumlah wisatawan yang dikelompokkan menjadi 5 terdiri dari *cluster_0* sebanyak 4 negara, yaitu Malaysia, Singapore, China, dan Timor Leste. *Cluster_1* sebanyak 4 negara, yaitu Philippines, Thailand, Taiwan, dan Saudi Arabia. *Cluster_2* sebanyak 11 negara, yaitu Brunei Darussalam, Vietnam, Myanmar/Burma, Hong Kong, Pakistan, Bangladesh, Srilanka, Nepal, Turki, Mesir (Egypt), dan Oman. *Cluster_3* sebanyak 25 negara. Dan *Cluster_4* sebanyak 3 negara, yaitu India, Japan, dan South Korea. Hasil analisis titik centroid kedua kluster ditunjukkan pada Tabel 9. Hasilnya menunjukkan bahwa *cluster_0* merupakan kluster terbanyak kunjungannya diikuti dengan *cluster*_4, *cluster*_1, *cluster*_2 dan *cluster*_3.

Tabel 9 Titik Centroid K-Medoids dengan k=5

Attribute	*Cluster_0*	*Cluster_1*	*Cluster_2*	*Cluster_3*	*Cluster_4*
2021	819488	2053	89	182	9497
2020	994590	31906	3612	428	75562
2019	1178381	157512	24051	2322	388316
2018	1762442	165912	25704	1941	358885
2017	960026	182086	18615	2328	423191

Dari perbandingan kedua metode klustering maka algoritma k-means klustering dengan nilai k=5 menjadi algoritma terbaik untuk mengelompokkan jumlah wisatawan yang datang ke Indonesia. Kelompok kluster dibagi menjadi 5 yaitu sangat banyak, banyak, sedang, sedikit, dan sangat sedikit. Negara dengan jumlah sangat banyak adalah sangat banyak adalah *cluster_2* yaitu Malaysia, Kategori banyak adalah *cluster_1* yaitu Singapore dan China, kategori sedang adalah *cluster_3* yaitu Timor Leste, kategori sedikit adalah *cluster_4* yaitu Philippines, India, Japan, South Korea, Taiwan, dan kategori sangat sedikit adalah *cluster_0* yaitu Brunei Darussalam, Thailand, Vietnam, Laos, Kamboja, Myanmar/Burma, Hong Kong, Pakistan, Bangladesh, Srilanka, Afganistan, Bhutan, Kazakhstan, Kirgistan, Korea Utara, Macao, Mongolia, Maladewa, Nepal, Tajikhistan, Turki, Turkmenistan, Uzbekistan, Saudi Arabia, Bahrain, Kuwait, Mesir (Egypt), Uni Emirat Arab, Yaman, Qatar, Irak, Iran, Israel, Jordan, Libanon, Oman, Palestina, dan Syiria.

4. Kesimpulan

Algoritma K-Means lebih baik dari pada Algoritma K-Medoid dalam kasus pengelompokan Negara berdasarkan jumlah kunjungan wisatawan mancanegara yang datang ke Indonesia. Hal ini dilihat dari nilai evaluasi *davies bouldin index*. Kluster terbaik dalam kasus ini adalah K-Means dengan nilai k=5 dengan hasil nilai *davies bouldin index* -0.302. Hasil pengelompokan negara dengan 5 kategori yang

ISSN: 2579-566X (Online)
ISSN: 2477-5274 (Print)

meliputi kategori sangat banyak, banyak, sedang, sedikit, dan sangat sedikit. Sehingga hasil pengelompokan tersebut dapat digunakan sebagai acuan pemerintah dalam menerapkan strategi untuk meningkatkan antusias wisatawan mancanegara datang ke Indonesia bagi Negara-negara yang jumlah wisatawannya masih rendah.

Referensi

Urbanus, N., & Febianti. (2017). Analisis dampak perkembangan pariwisata terhadap perilaku konsumtif masyarakat wilayah bali selatan. *Jurnal Kepariwisataan Dan Hospitalitas*, *1*(No.2), 118–133.

Irfiani, E., Sulistia Rani, S., Kamal Raya No, J., Road Barat Cengkareng Jakarta Barat, R., Nusa Mandiri Jl Kramat Raya No, S., & Pusat, J. (2018). Algoritma K-Means *Clustering* untuk Menentukan Nilai Gizi Balita. *Jurnal Sistem Dan Teknologi Informasi*, *6*(4), 161–168.

Alfatah, D. (2021). Application of the K-Means *Clustering* Algorithm in Mapping the Regional Voter Strategy for the Legislative Candidates for the DPR RI Penerapan Algoritma K-Means *Clustering* dalam Memetakan Strategi Daerah Pemilih pada Calon Legislatif DPR RI. *Jurnal Komitek*, *1*(2), 435–443.

Farokhah, L., & Aditya, R. (2017). Implementasi K-Means Klustering Untuk Rekomendasi Tema Tugas Akhir Pada Stmik Asia Malang. *Jurnal Teknologi Dan Manajemen Informatika*, *3*(2), 142–148. https://doi.org/10.26905/jtmi.v3i2.1329

Mustofa, Z., & Iman Saufik Suasana. (2020). Algoritma *Clustering* K-Medoids Pada E-Government Bidang Information and Communication Technology Dalam Penentuan Status Edgi. *Jurnal Teknologi Informasi Dan Komunikasi*, *9*(1), 1–10. https://doi.org/10.51903/jtikp.v9i1.162

Riyanto, B. (2019). Penerapan Algoritma K-Medoids *Clustering* Untuk Pengelompokkan Penyebaran Diare Di Kota Medan (Studi Kasus: Kantor Dinas Kesehatan Kota Medan). *KOMIK (Konferensi Nasional Teknologi Informasi Dan Komputer)*, *3*(1), 562–568. https://doi.org/10.30865/komik.v3i1.1659

Alodia, D. A., Fialine, A. P., Endriani, D., & Widodo, E. (2021). Implementasi Metode K-Medoids *Clustering* untuk Pengelompokan Provinsi di Indonesia Berdasarkan Indikator Pendidikan. *Sepren*, *2*(2), 1–13. https://doi.org/10.36655/sepren.v2i2.606

Dwilestari, G., Mulyawan, Martanto, & Ali, I. (2021). Analisis *Clustering* menggunakan K-Medoid pada Data Penduduk Miskin Indonesia. *JURSIMA: Jurnal Sistem Informasi Dan Manajemen*, *9*(3), 282–290.

Jannah, S. M. (2022). *BPS Catat Kunjungan Wisman Sepanjang 2021 Hanya 1,56 Juta Orang*. Tirto.Id. https://tirto.id/bps-catat-kunjungan-wisman-sepanjang-2021-hanya-156-juta-orang-gosd

Dinata, R. K., Safwandi, S., Hasdyna, N., & Azizah, N. (2020). Analisis K-Means *Clustering* pada Data Sepeda Motor. INFORMAL: Informatics Journal, 5(1), 10. https://doi.org/10.19184/isj.v5i1.17071

Badan Pusat Statistik. (n.d.). Retrieved October 1, 2022, from https://www.bps.go.id/indicator/16/1470/1/kunjungan-wisatawan-mancanegara-per-bulan-menurut-kebangsaan.html

Muningsih, E., Maryani, I., & Handayani, V. R. (2021). Penerapan Metode K-Means dan Optimasi Jumlah *Cluster* dengan Index Davies Bouldin untuk *Clustering* Propinsi Berdasarkan Potensi Desa. *Jurnal Sains Dan Manajemen*, *9*(1), 95–100. https://ejournal.bsi.ac.id/ejurnal/index.php/evolusi/article/view/10428/4839