Implementasi Data Mining Untuk Clustering Asal Daerah Pengunjung Pada Wisata Alam Puncak Sempur Kabupaten Karawang

Adinda Ryanda
Universitas Buana Perjuangan Karawang
Karawang, Indonesia
If18.adindaryandat@mhs.ubpkarawang.ac.id

Ahmad Fauzi
Universitas Buana Perjuangan Karawang
Karawang, Indonesia
Afauzi@ubpkarawang.ac.id

Elsa Elvira Awal
Universitas Buana Perjuangan Karawang
Karawang, Indonesia
elsaelvira@ubpkarawang.ac.id

Abstract*-- Setiap pengelola objek wisata melakukan pendataan kunjungan wisatawan, pendataan ini sangat diperlukan dalam menentukan target promosi. Puncak sempur merupakan salah satu objek wisata alam di Karawang yang melakukan pendataan kunjungan wisatawan melalui buku pengunjung, namun tidak mudah dalam menentukan daerah kurang peminat jika dilihat dari buku daftar pengunjung. Teknik yang tepat dalam permasalahan tersebut berdasarkan penelitian sebelumnya adalah teknik Data Mining *clustering*, yang dimana prosesnya melalui pengelompokan dengan proses algoritma K-means dan K-medoids. Variabel yang dibutuhkan dalam *data mining* ini berupa asal daerah pengunjung dan jumlah pengunjung, yang di mana data tersebut dapat ditemui pada buku daftar pengunjung. Data yang didapat terdiri dari 1.797 wisatawan dan 32 daerah akan dihitung manual dan bahasa phyton menggunakan algoritma k–means dan dibandingkan dengan perhitungan algoritma k-medoid, hasil perhitungan algoritma k-means dan k-medoid berbeda sehingga dilakukan evaluasi untuk mengetahui tingkat akurasi dengan menggunakan metode *Silhouette coefficient*. Hasil evaluasi menunjukan algoritma k-means lebih baik akurasinya yaitu dengan nilai 0,8% sedangkan algoritma k-medoid memiliki nilai akurasi di bawah algoritma k-means yaitu 0,7%. Jadi hasil perhitungan algoritma k-means yang lebih baik untuk hasil penelitian ini yaitu 1 daerah asal wisatawan yang tergolong ramai (C0), 1 daerah asal wisatawan yang tergolong sedang (C1), dan 28 daerah asal wisatawan yang tergolong sepi (C2).

Kata Kunci: *Clustering, Data mining*, k-means, k-medoid, puncak sempur.

I. PENDAHULUAN

Keberagaman objek wisata yang ada di Indonesia dan tersebar luas di setiap daerah nya menjadikan objek wisata salah satu ciri khas suatu daerah. Banyak jenis objek wisata yang ada di Indonesia di antaranya wisata bahari, wisata budaya, wisata pertanian, wisata sejarah atau ziarah, wisata alam dan wisata konvensi. Wisata tersebut mempunyai daya tarik masing-masing dalam mengundang para wisatawan. Wisata alam dan wisata sejarah menjadi daya tarik untuk wisatawan saat ini, *back to nature* adalah jenis wisata yang sedang maraknya pada masa sekarang. Suasana yang natural yang sangat jauh berbeda dengan suasana kehidupan pada setiap harinya, membuat objek wisata alam sangat diburu oleh wisatawan belakangan ini.

Sektor pariwisata merupakan sektor yang berperan penting dalam berbagai aspek, mulai dari perekonomian rakyat, pembangunan negara hingga penyembuh suntuk nya seseorang dalam kehidupan sehari-hari nya. Namun tidak semua pariwisata dapat memuaskan pengunjung di karenakan beberapa aspek, di antaranya fasilitas yang disediakan dan peningkatan fasilitas yang mengikuti jaman (modernisasi) [1]. Fasilitas yang disediakan oleh pengelola wisata alam puncak sempur dan suasana yang mengikuti jaman menjadi salah satu faktor dalam kunjungan wisata, namun tidak semua wisatawan mengetahuinya hal itu yang menyebabkan sedikitnya pengunjung. Sehingga adanya pengelompokan pengunjung objek wisata alam puncak sempur berdasarkan asal daerah guna mengetahui tingkat tersebarnya objek wisata alam puncak sempur ini, pengelompokannya terdiri dari ramai pengunjung, sedang pengunjung, dan sepi pengunjung.

Penelitian terkait pengelompokan, pertama penelitian untuk mengelompokkan objek wisata unggulan di DKI Jakarta menggunakan metode K-means yang dilakukan oleh Maulida, data yang didapatkan untuk penelitian ini dari BPS (Badan Pusat Statistik) berupa 8 objek wisata unggulan di DKI Jakarta. Tahapan metode data mining yang digunakan dituangkan dalam bentuk flowchart, masuk ke dalam perhitungan K-means yang di mana jumlah cluster dibagi menjadi 3 yaitu C1, C2 dan C3 setelah itu menetukan data mana yang paling dekat dengan nilai cluster yang ada. Hasil dari penelitian ini yaitu diketahuinya objek wisata yang ramai, sedang dan tidak ramai, dari hasil ini maka bisa menjadi catatan bagi Pemerintah DKI Jakarta untuk objek wisata yang masuk dalam kategori C3 atau Sepi Pengunjung[2].

Penelitian selanjutnya yaitu pengelompokan kunjungan wisatawan pada objek wisata di Provinsi Jambi yang dilakukan oleh Novhirtamely, Lailyn dan Mona, yang di mana penelitian ini bertujuan agar mengetahui wisata unggulan yang ada di Provinsi Jambi. Data yang digunakan 174 tempat wisata, didapatkan 10 wisata unggulan pada provinsi Jambi. Pada penelitian ini pun tidak sekedar mengetahui wisata unggulan di provinsi Jambi namun ada pula pembuatan aplikasi dengan menampilkan hasil clustering tersebut[3]. Penelitian berikutnya yaitu mengelompokan lapangan kerja informal sektor non-pertanian yang ditulis oleh Khairunnissa Fanny Irnanda, Agus Perdana Windarto, Dedy Hartama, dan Anjar Wanto. Penelitiannya mengambil sumber data dari BPS yang terdiri dari 34 wilayah yang ada di Indonesia yang diolah menggunakan metode data mining clustering dengan algoritma k-medoid dan dituangkan dalam bentuk flowchart. Pengelompokan dibagi menjadi 2 yaitu cluster tingkat lapangan kerja rendah

dan tingkat lapangan kerja tinggi, hasil dari C1 lah yang akan dijadikan pusat perhatian bagi pemerintah untuk menambah lapangan kerja informal pada sektor non-pertanian[4].

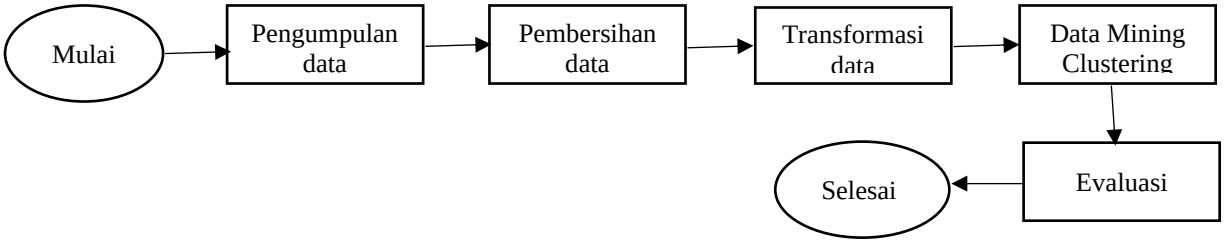
II. METODOLOGI PENELITIAN

A. *Objek Penelitian*

Objek Penelitian ini yaitu jumlah kunjungan wisatawan berdasarkan asal daerah di objek wisata alam Puncak Sempur Kabupaten Karawang tahun 2021. Data yang didapatkan berupa hasil wawancara yang akan dikonversikan dalam bentuk tabel data, agar memudahkan dalam pengaplikasiannya ke dalam *data mining*. Setelah dikonversikan, data tersebut akan diproses dengan menggunakan metode *clustering data mining* dengan algoritma K-Means dan K-Medoid. Sampai nantinya akan ditemukan daerah mana saja yang dominan mengunjungi objek wisata alam Puncak Sempur, pengunjung lokal atau pengunjung luar daerah.

B. *Prosedur Penelitian*

Penelitian ini Mengikuti proses *Knowledge Discovery in Database* (KDD) yaitu nama lain dari data mining, namun pada proses KDD sendiri ada proses data mining. Alur penelitian yang dilakukan yaitu dimulai dari pengumpulan data, analisis data, *clustering data mining* yang terakhir pengujian dan mendapatkan hasil yang di inginkan. Berikut diagram prosedur penelitian :



Gambar 1 Tahapan Prosedur Penelitian

C. *Pengumpulan Data*

Tahap pertama dari penelitian ini yaitu mengumpulkan data dari sumber data atau objek yang ditentukan. Tahapan ini terdiri dari 3 metode yaitu wawancara, Observasi dan Studi kepustakaan. Hasil yang didapatkan dari wawancara yaitu mengetahuinya jumlah pengunjung yang datang pada tahun 2021 melalui buku daftar pengunjung, Dari Observasi sendiri peneliti dapat mengetahui kondisi wisata alam secara langsung terutama dalam fasilitasnya, dan Studi kepustakaan yang dibaca membantu dalam penyusunan penulisan penelitian ini. Pengumpulan data ini dari ke tiga metode di atas adalah jumlah pengunjung sebanyak 2.172 pada tahun 2021 dengan 38 daerah yang terdiri dari luar Karawang dan dalam Karawang, Fasilitas yang disediakan pun sudah layak pakai.

D. *Pembersihan Data*

Banyak data yang tidak diperlukan saat dilakukannya pengumpulan data, dengan cara memisahkannya dengan manual menggunakan Microsoft Excel 2013 antara data yang dibutuhkan atau tidak untuk tahap selanjutnya. Data dibersihkan dalam bentuk data sampah, kurang nya pengejaan sampai dengan melengkapi data yang kurang lengkap. Data yang diperlukan dalam penelitian ini hanyalah jumlah kunjungan wisata dan asal daerah Kabupaten Karawang, sehingga data lainnya dapat di hapus atau di bersihkan agar data menjadi efektif saat proses perhitungan.

E. *Tranformasi Data*

Ubah atau gabungkan data menjadi dalam format yang sesuai pemrosesan dalam data mining. Sebagian dari metode data mining butuh format data yang khusus untuk dapat di terapkan, seperti metode asosisasi dan clustering hanya dapat menerima input data klasifikasi, oleh karena itu data dalam bentuk bilangan berurutan perlu di bagi menjadi beberapa selang waktu dan proses ini yang disebut dengan Transformasi Data[5]. Mengubah atau merombak data sehingga data sesuai, fitur ditransformasi berguna untuk menyajikan data tergantung tujuan yang ingin dicapai dan prosesnya ubah data yang dipilih agar sesuai dengan data dengan proses data mining.

F. *Implementasi Data Mining*

1. Proses clustering Algoritma K-means

Input data yang berupa jumlah kunjungan wisatawan di objek wisata puncak sempur. Proses kedua buat 3 clustering terlebih dahulu dengan simbol C1(Ramai pengunjung), C2(Sedang pengunjung), dan C3(Sepi pengunjung). Proses ketiga menentukan pusat kelompok, titik pusat / centroid yang di tentukan secara acak. Proses keempat menghitung jarak terdekat dari masing-masing data ke setiap centroid/titik pusat, dengan rumus :

$$D_e = \sqrt{(x_i - S_I)^2} + (y_i - t_i)^2 \qquad (1)$$

berdasarkan persamaan 1 di atas De adalah Euclidean Distance, xi merupakan data kriteria dan μj merupakan centroid pada cluster ke-j. Langkah ke empat kelompokan data berdasarkan kedekatan dengan centroid. jika sudah terbentuk data baru dari iterasi 1 maka hitung kembali centroid baru dengan cara menghitung nilai rata-rata disetiap anggota centroid, dengan rumus :

$$x_j(\mathrm{t}\ +1) = \frac{1}{N_{sj}} \sum_{j \in sj} x_j \qquad (2)$$

keterangan persamaan 2 yaitu μj(t+1) merupakan centroid baru pada iterasi ke (t+1) dan Nsj merupakan banyaknya data pada cluster saja. Lakukan langkah 3 sampai 4 sampai anggota tiap cluster tidak ada yang berubah [6]. Lalu buat iterasi 2 dengan menghitung kembali jarak terdekat masing-masing data ke setiap centroid atau titik pusat. jika hasil cluster sama atau tidak berubah maka proses dihentikan atau selesai, namun jika berubah hasil cluster maka lakukan kembali proses mulai dari menghitung jarak terdekat data ke setiap centroid. Output nya akan terlihat berupa 3 kelompok yang terkategorikan Ramai pengunjung, Sepi pengunjung, dan Sedang pengunjung.

2. Proses Clustering Algoritma K-Medoid

Input data yang berupa jumlah kunjungan wisatawan di objek wisata alam puncak sempur. Kedua pembuatan pusat medoid, penelitian ini menggunakan 3 cluster yaitu C1, C2 dan C3. Ketiga pilih masing-masing centroid sebagai kandidat medoid baru. Keempat hitung jarak antar objek dengan masing-masing centroid, dengan rumus :

$$\mathrm{d} = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \qquad (3)$$

Kelima hitung total simpangan (S) dengan menghitung nilai total distance baru – total distance lama. Jika S < 0, maka tukar objek dengan data cluster untuk membentuk sekumpulan k objek baru sebagai medoid. Lakukan point ke 4 dan ke 5 hingga hasil cluster tidak berubah, maka akan didapatkan output berupa 3 kelompok yaitu Ramai pengunjung, Sepi pengunjung, dan Sedang pengunjung.

G. *Evaluasi Silhoutte Coefficient*

Evaluasi dalam clustering dapat dihitung tingkat ke akurasiannya menggunakan Silhouette Coefficient, metode evaluasi yang terdiri dari penggabungan dua metode, metode pertama adalah Cohesion yang berfungsi untuk menghitung jarak antar objek pada setiap cluster dan metode kedua adalah metode Separation yang berfungsi menghitung jarak antara cluster satu dengan cluster lainnya. Metode evaluasi ini memiliki nilai dari -1 hingga 1 yang dimana jika hasil lebih mendekati nilai 1 maka metode clustering lebih bagus juga dan sebaliknya jika hasil mendekati nilai -1 maka metode clustering sangat buruk[7]. Metode Silhouette Coefficient ini dapat di terapkan dengan persamaan berikut :

$$\mathrm{sil}(c) = \mathrm{sil}(k)\frac{1}{|k|}\sum_{i=1}^{k} sil(c_i) \qquad (4)$$

Keterangan dari persamaan 4 di atas yaitu sil(k) adalah nilai Silhouette semua cluster, |k| adalah banyaknya cluster, dan sil(ci) adalah rata-rata nilai Silhouette.

III. HASIL DAN PEMBAHASAN

A. Hasil Pengumpulan Data

Data yang didapatkan dari observasi langsung ke tempat wisata alam puncak sempur yaitu data dari buku daftar pengunjung pada tahun 2021. Jumlah pengunjung asal daerah kabupaten karawang mencapai 1.797 jiwa, dengan kategori bermalam (camping) dan tidak bermalam. Data yang didapat berupa buku yang diubah menjadi bentuk excel, berikut data pengunjung wisata alam puncak sempur kabupaten Karawang 2021 :

NO	TANGGAL	NAMA KETUA	JUMLAH ANGGOTA	ALAMAT	TUJUAN		Keterangan
					CAMP	TIDAK CAMP	
1	01-Januari-2021	Pungki	120 orang	LOJI	✔		Archiless Goes
2	02-Januari-2021	ucay	3 orang	CACABAN	✔		
3	02-Januari-2021	adzuar kumar	2 orang	CIDORO		✔	
4	04-Januari-2021	ahmad julpa	5 orang	BADAMI	✔		
5	04-Januari-2021	ilmi Nurhidayah	5 orang	BOGOR	✔		
...	...	...	...	...	...	...	...
190	31-Desember-2021	sidiq	4 orang	KARANG PAWITAN	✔		

Gambar 2 Hasil Pengumpulan Data

B. *Hasil Pembersihan Data*

Variabel yang diperlukan dalam penelitian ini yaitu asal daerah pengunjung dan jumlah pengunjun. Oleh karena itu data mentah yang didapatkan dari buku daftar pengunjung perlu dibersihkan dari variabel-variabel yang tidak dibutuhkan. Adapun data yang dibersihkan seperti tanggal, camping/tidak camping, pengunjung luar kabupaten Karawang dan keterangan yang dimana tidak diperlukan dalam perhitungan sehingga dapat dihapus.

C. *Hasil Transformasi Data*

Transformasi data yang dihasilkan dari data kunjungan wisata puncak sempur yaitu perubahan format agar dapat terbaca oleh perhitungan manual maupun menggunakan bahasa phyton untuk clustering. Tranformasi ini menghasilkan kelompok kecamatan yang ada di Kabupaten Karawang dengan 32 daerah asal Karawang. Nama ketua yang dihilangkan dan juga penjumlahan wisatawan berdasarkan bulan dan daerah selama tahun 2021. Transformasi ini menambahkan seluruh kecamatan yang ada di kabupaten Karawang yang berjumlah 30 kecamatan. Hasil Transformasi Data sebagai berikut :

KECAMATAN	Januari	Februari	Maret	April	Mei	Juni	Juli	Agustus	September	Oktober	November	Desember
Banyusari	0	0	0	0	0	0	0	0	0	0	0	0
Batujaya	0	0	0	0	0	0	12	0	10	15	0	0
Ciampel	0	0	0	5	0	0	0	4	0	0	0	2
Cibuaya	0	0	0	0	0	0	0	0	0	0	0	0
Cikampek	0	0	0	0	0	230	0	0	12	4	0	0
Cilamaya Kulon	0	0	0	0	0	0	0	0	0	0	0	0
Cilamaya Wetan	0	0	0	0	0	0	0	0	0	0	0	0
Cilebar	0	0	0	0	0	0	0	0	0	0	0	0
Jatisari	0	0	0	0	0	0	0	10	0	6	0	16
Jayakerta	0	0	0	0	0	0	0	0	0	0	0	0
Karawang Barat	0	0	0	0	0	0	0	4	10	0	0	4
Karawang Timur	0	0	0	5	0	0	0	5	10	0	0	6
Klari	0	0	4	0	0	0	0	14	11	0	3	40
Kotabaru	0	0	0	0	0	0	0	65	0	5	0	0
Kutawaluya	0	0	0	0	0	0	0	0	0	0	0	0
Lemahabang	0	0	0	0	0	0	0	0	0	0	0	0
Majalaya	0	0	0	0	0	0	0	0	0	0	0	0
Pakisjaya	0	0	0	0	0	0	0	0	0	0	0	0
Pangkalan	0	0	0	0	0	0	0	0	0	0	0	0
Pedes	0	0	0	0	0	0	4	1	4	12	0	0
Purwasari	4	0	0	0	0	0	0	3	9	0	0	5
Rawamerta	0	0	0	0	0	5	0	0	0	3	0	0
Rengasdengklok	0	0	0	0	0	0	2	152	20	7	144	0
Tegalwaru	127	0	7	4	10	2	0	41	167	4	8	178
Talagasari	0	0	0	0	9	0	0	1	0	0	0	0
Telukjambe Barat	5	0	5	0	0	3	0	17	20	0	0	4
Telukjambe Timur	0	6	6	2	16	7	16	25	20	120	62	0
Tempuran	0	0	1	0	0	2	0	0	11	4	0	0
Tirtajaya	0	0	0	0	0	0	0	0	0	0	0	0
Tirtamulya	0	0	0	0	0	0	0	0	0	0	0	0

Gambar 3 Hasil Transformasi Data

D. *Hasil Klasterisasi Kunjungan Wisatawan*

1. Perhitungan k-means
   a. Perhitungan Manual

Pada perhitungan manual ini yang menggunakan Microsoft Excel 2013, dengan iterasi sebanyak dua kali sampai di temukannya hasil akhir yaitu C1 = 1 daerah tinggi peminat, C2 = 1 daerah sedang peminat, dan C3 = 28 daerah rendah peminat yang di buktikan pada perhitungan iterasi kedua atau terakhir berikut :

iterasi Kedua				Hasil	Hasil	Hasil
data ke-i	c1	c2	c3	Cluster 1	Cluster 2	Cluster 3
1	278.63	230.35	15.49			cluster 3
2	273.21	230.58	20.15			cluster 3
3	276.76	230.45	13.55			cluster 3
4	278.63	230.35	15.49			cluster 3
5	354.59	0.00	229.92		cluster 2	
6	278.63	230.35	15.49			cluster 3
7	278.63	230.35	15.49			cluster 3
8	278.63	230.35	15.49			cluster 3
9	267.32	231.09	15.97			cluster 3
10	278.63	230.35	15.49			cluster 3
11	269.58	230.11	13.21			cluster 3
12	268.13	230.23	13.85			cluster 3
13	245.78	233.96	38.90			cluster 3

14	276.61	239.31	55.05			cluster 3
15	278.63	230.35	15.49			cluster 3
16	278.63	230.35	15.49			cluster 3
17	278.63	230.35	15.49			cluster 3
18	278.63	230.35	15.49			cluster 3
19	278.63	230.35	15.49			cluster 3
20	276.22	230.32	14.23			cluster 3
21	267.80	230.16	14.00			cluster 3
22	278.61	225.32	15.19			cluster 3
23	316.89	311.15	197.09			cluster 3
24	0.00	354.59	272.47	cluster 1		
25	278.30	230.53	16.86			cluster 3
26	259.90	227.96	20.62			cluster 3
27	293.95	261.35	130.16			cluster 3
28	272.11	228.00	15.19			cluster 3
29	278.63	230.35	15.49			cluster 3
30	278.63	230.35	15.49			cluster 3
Jumlah Anggota Cluster				1	1	28

Gambar 4 Hasil Algoritma K-means

b. Perhitungan dengan phyton

Perhitungan Phyton ini dilakukan guna memperkuat hasil dari perhitungan manual sebelumnya, hasil perhitungan phyton ini sama dengan perhitungan manual yaitu : C1 = 1 daerah tinggi peminat, C2 = 1 daerah sedang peminat dan C3 = 28 daerah rendah peminat. Adapun source code phyton perhitungan k-means beserta hasilnya sebagai berikut :

```
jumlah anggota setiap cluster
cluster 0 =  1
cluster 1 =  1
cluster 2 =  28
```

Gambar 5 Hasil Perhitungan Phyton K-means

2. Perhitungan K-medoids

a. Perhitungan Manual

Berdasarkan urutan perhitungan manual yang sudah disebutkan sebelum nya, bahwa nilai medoid sementara yang akan di ambil menjadi nilai hasil jika total simpangan lebih besar dari 0 (0<), maka hasil dari perhitungan manual k-medoids yaitu C1 = 2 daerah tinggi peminat, C2 = 1 daerah sedang peminat, dan C3 = 27 daerah rendah peminat. Berikut hasil dari perhitungan medoid sementara yang dijadikan hasil akhir :

MEDOID SEMENTARA				Kedekatan	Hasil Cluster 1	Hasil Cluster 2	Hasil Cluster 3
data ke-i	c1	c2	c3				
1	141.09	230.35	9.06	9.06			cluster 3
2	126.46	230.58	23.47	23.47			cluster 3
3	140.47	230.45	10.91	10.91			cluster 3
4	141.09	230.35	9.06	9.06			cluster 3
5	261.35	0.00	230.53	0.00		cluster 2	
6	141.09	230.35	9.06	9.06			cluster 3
7	141.09	230.35	9.06	9.06			cluster 3

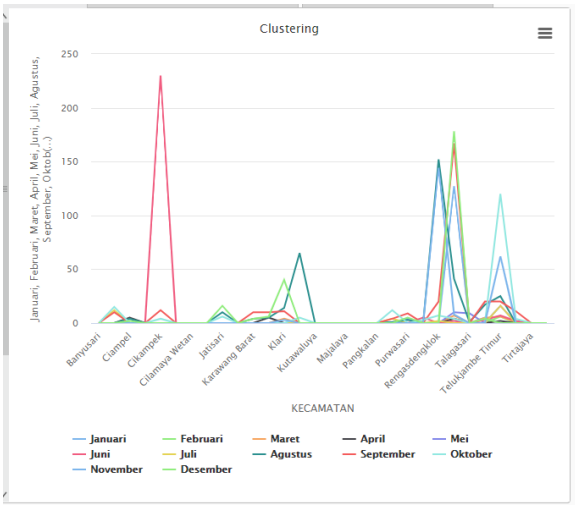
ISSN: 2715-2766
Vol. IV No: 2, Juli 2023

8	141.09	230.35	9.06	9.06			cluster 3
9	135.49	231.09	21.31	21.31			cluster 3
10	141.09	230.35	9.06	9.06			cluster 3
11	139.42	230.11	14.35	14.35			cluster 3
12	139.36	230.23	16.06	16.06			cluster 3
13	142.44	233.96	44.68	44.68			cluster 3
14	140.38	239.31	64.82	64.82			cluster 3
15	141.09	230.35	9.06	9.06			cluster 3
16	141.09	230.35	9.06	9.06			cluster 3
17	141.09	230.35	9.06	9.06			cluster 3
18	141.09	230.35	9.06	9.06			cluster 3
19	141.09	230.35	9.06	9.06			cluster 3
20	129.87	230.32	16.03	16.03			cluster 3
21	139.74	230.16	14.39	14.39			cluster 3
22	138.38	225.32	10.77	10.77			cluster 3
23	190.26	311.15	209.93	190.26	cluster 1		
24	293.95	354.59	278.30	278.30			cluster 3
25	140.18	230.53	0.00	0.00			cluster 3
26	137.54	227.96	28.50	28.50			cluster 3
27	0.00	261.35	140.18	0.00	cluster 1		
28	136.41	228.00	14.97	14.97			cluster 3
29	141.09	230.35	9.06	9.06			cluster 3
30	141.09	230.35	9.06	9.06			cluster 3
Jumlah Kedekatan				866.54	2	1	27

Gambar 6 Hasil Algoritma K-medoids

b. Perhitungan dengan Rapidminer

Pada perhitungan k-medoids menggunakan Rapidminer hasilnya sama dengan perhitungan manual sebelumnya yaitu C1 = 2 daerah tinggi peminat, C2 = 1 daerah rendah peminat dan C3 = 27 daerah rendah peminat. Dibuktikan dengan hasil clustering visualisasi berikut :



Gambar 7 Hasil Visualisasi Rapidminer K-medoids

E. Hasil Evaluasi Silhouette Coefficient

Hasil yang didapat dari setiap perhitungan algoritma k-means dan K-Medoids dapat digunakan untuk menentukan tingkat akurasi hasil, dimana akan menghasilkan kesimpulan metode clustering manakah yang cocok dalam pengelompokankunjungan wisata ini dengan menggunakan metode Silhouette coefficient. Perhitungan Silhouette untuk metode K-means bernilai 0,8% dan metode K-Medoids bernilai 0,7% yang dimana nilai Silhouette k-means lebih mendekati nilai 1 daripada nilai Silhouette K-Medoids. Pembahasan dari perhitungan di atas adalah untuk perhitungan k-means yang menggunakan metode manual dan metode phyton dengan jupyter lab memiliki hasil clustering yang sama yaitu untuk C0 = 1 daerah, C1 = 1 daerah, dan C2 = 28 daerah. Sedangkan perhitungan K-Medoids yang juga menggunakan perhitungan manual dan phyton memiliki hasil clustering yang berbeda dengan k-means yaitu dengan hasil C0 = 2 daerah, C1 = 1 daerah, dan C2 = 27 daerah. Hasil clustering yang berbeda itu disebabkan oleh beda nya pemilihan centroid baru antar algoritma yang dimana algoritma k-means menentukan centroid baru dengan mencari rata-rata dari anggota cluster iterasi sebelumnya, sedangakan algoritma K-Medoids menentukan centroid baru dengan mengambil secara acak kembali dari dataset.

IV. KESIMPULAN DAN SARAN

A. *Kesimpulan*

Hasil pengimplementasian data mining clustering algoritma K-means untuk menentukan daerah yang kurang mengunjungi wisata puncak sempur dan dihitung secara manual dan menggunakan phyton yaitu C1 = 1 daerah yang tinggi peminat, C2 = 1 daerah yang sedang peminat, C3 = 28 daerah rendah peminat. Sedangkan hasil pengimplementasian algoritma K-Medoids yaitu C1 = 2 daerah yang tinggi peminat, C2 = 1 daerah sedang peminat, C3 = 27 daerah tergolong rendah peminat. Hasil evaluasi untuk algoritma k-means bernilai 0,8% dan hasil evaluasi untuk algoritma K-Medoids bernilai 0,7% , yang dapat disimpulkan bahwa akurasi hasil clustering k-means lebih baik daripada metode K-Medoids. Dapat disimpulkan bahwa perhitungan k-means lebih baik daripada K-Medoids yang berarti masih ada 28 kecamatan di kabupaten Karawang yang kurang pengunjung. Bahkan masih ada Kecamatan di Karawang yang belum sama sekali mengunjungi wisata alam puncak sempur Karawang.

B. *Saran*

1. Untuk pihak pengelola, diperlukannya promosi yang lebih semisal promosi lewat media sosial yang harus aktif dalam memberi info tentang wisata alam puncak sempur Karawang. Hal tersebut dapat berpengaruh terhadap wisatawan luar daerah yang membaca informasi tersebut dan ingin mengunjungi wisata alam puncak sempur Karawang.
2. Untuk Disparbud, agar mengumpulkan dan menyampaikan ke setiap kepala camat yang daerah nya belum sama sekali mengunjungi wisata alam puncak sempur supaya mengenalkan wisata alam ini untuk dikunjungi.
3. Untuk pengembangan penelitian selanjutnya, penelitian selanjutnya dapat menggunakan perbandingan algoritma lain.

PENGAKUAN

Naskah ilmiah ini adalah sebagian dari penelitian Tugas Akhir milik Adinda Ryanda, dengan judul Clustering Asal Daerah Pengunjung pada Wisata Alam Puncak Sempur Karawang berdasarkan Kecamatan menggunakan Algoritma K-means dan K-medoids yang dibimbing oleh Ahmad Fauzi dan Elsa Elvira Awal.

DAFTAR PUSTAKA

[1] Van Daffa, K., & Ratnasari, I. (2022). Pengaruh Citra Destinasi dan Promosi Terhadap Kepuasan Wisatawan di Puncak Sempur Kabupaten Karawang. *Jurnal Ilmiah Wahana Pendidikan*, *8*(1), 362-372.

[2] Maulida, L. (2018). Penerapan *Data mining* dalam Mengelompokkan Kunjungan Wisatawan ke Objek Wisata Unggulan di Prov. DKI Jakarta dengan K-Means. *JISKA (Jurnal Informatika Sunan Kalijaga)*, *2*(3), 167-174.

[3] Kahar, N., Puad, L., & Istuti, M. (2020). ANALISIS PENGELOMPOKAN KUNJUNGAN WISATAWAN KE OBJEK WISATA UNGGULAN DI PROVINSI JAMBI BERBASIS *DATA MINING*. *JURNAL AKADEMIKA*, *13*(1), 66-72.

[4] Irnanda, K. F., Windarto, A. P., Hartama, D., & Wanto, A. (2019). ANALISA METODE *DATA MINING* PADA PENGELOMPOKAN LAPANGAN KERJA INFORMAL SEKTOR NON-PERTANIAN. *KOMIK (Konferensi Nasional Teknologi Informasi dan Komputer)*, *3*(1).

[5] Ali, I., Dikananda, A. R., Ma'ruf, F. A., & Abdurohman, M. (2021). PENGELOMPOKAN JUMLAH PENDUDUK BERDASARKAN KATEGORI USIA 0-18 TAHUN DENGAN MENGGUNAKAN ALGORITMA K-MEANS UNTUK MENENTUKAN PENGEMBANGAN POTENSI DESA WISATA DI KABUPATEN CIREBON. *JURNAL MANAJEMEN INFORMATIKA (JUMIKA)*, *8*(1).

[6] Indriyani, F., & Irfiani, E. (2019). *Clustering* Data Penjualan pada Toko Perlengkapan Outdoor Menggunakan Metode K-Means. *JUITA: Jurnal Informatika*, *7*(2), 109-113.

[7] Paembonan, S., & Abduh, H. (2021). Penerapan Metode *Silhouette* Coefficient untuk Evaluasi *Clustering* Obat. *PENA TEKNIK: Jurnal Ilmiah Ilmu-Ilmu Teknik*, *6*(2), 48-54.