

Geo-Location Clustering Untuk Menentukan Tempat Tujuan Wisata Dengan Menggunakan Algoritma PAM (Partitioning Around Medoids)

Risna Agustin¹, Asep Id Hadiana², Fajri Rahmat Umbara³

*Program Studi Informatika
Universitas Jenderal Achmad Yani
Jl. Terusan Jend. Sudirman, Cimahi, Jawa Barat Indonesia*

¹risnaagustin18@if.unjani.ac.id, ²asep.hadiana@lecture.unjani.ac.id, ³fajri.rakhmat@lecture.unjani.ac.id

Abstrak

Pariwisata adalah suatu kegiatan perjalanan dimana seorang individu atau kelompok mengunjungi suatu tempat tertentu untuk rekreasi atau mempelajari keunikan daya tarik tempat wisata yang dikunjungi. Salah satu tempat wisata di Indonesia yang banyak dikunjungi adalah Pulau Jawa dengan total kunjungan mencapai 186.894.200 jiwa dalam kurun waktu 5 tahun dimulai dari tahun 2014 hingga 2018. Dalam melakukan sebuah kunjungan wisata biasanya wisatawan akan melakukan proses perencanaan perjalanan dengan melihat informasi yang ada di internet, kemudian wisatawan akan menuliskan tempat tujuan yang akan dikunjungi. Proses merencanakan perjalanan wisata ini dapat membantu dalam mengatur perjalanan sehingga dapat mengunjungi setiap tempat wisata yang tersedia. Oleh karena itu dibutuhkan pengelompokan data tempat wisata untuk menentukan tempat wisata yang akan dikunjungi. Dalam penelitian ini algoritma PAM (Partitioning Around Medoids) digunakan untuk melakukan proses geo-location clustering pada data tempat wisata. Nilai K yang digunakan di inisialisasikan sebagai jumlah hari kunjungan. Penelitian ini menghasilkan 3 cluster, dimana cluster 0 memiliki jumlah kunjungan wisata sebanyak 284 tempat wisata dengan rata-rata koordinat masuk kedalam wilayah Jawa Barat, cluster 1 memiliki jumlah kunjungan wisata sebanyak 470 tempat wisata dengan rata-rata koordinat masuk kedalam wilayah Jawa Tengah, dan cluster 2 memiliki jumlah kunjungan wisata sebanyak 260 tempat wisata dengan rata-rata koordinat masuk kedalam wilayah Jawa Timur. Hasil evaluasi menggunakan metode Silhouette Coefficient untuk nilai K=3 yaitu 0,3386 sehingga hasil cluster yang terbentuk cukup baik.

Keywords: Geo-Location Clustering, PAM, Tempat Wisata.

I. PENDAHULUAN

Data *mining* adalah proses mengidentifikasi pola dan tren data untuk mendapatkan informasi yang berguna dalam kumpulan data yang sangat besar[1]. Algoritma PAM merupakan salah satu algoritma dalam teknik *clustering* yang ada pada data *mining*, algoritma ini akan melakukan klasterisasi dataset objek n ke dalam k *cluster* yang dikenal sebagai *apriori*[2]. Teknik *clustering* ini dapat digunakan dalam pengelompokan sebuah data lokasi atau disebut dengan *geo-location clustering*, dengan mengumpulkan data berupa data spasial atau data yang dapat menunjukkan lokasi di permukaan bumi. Dengan adanya data lokasi maka teknik GIS

(*Geographic Information System*) dapat digunakan sehingga dapat membantu dalam memahami pola, hubungan, dan konteks dalam geografis[3].

Dalam prosesnya data *mining* sangat membantu untuk mengelola data banyak, salah satunya dalam mengelola data pada bidang pariwisata. Pariwisata adalah suatu kegiatan perjalanan dimana seorang individu atau kelompok mengunjungi suatu tempat tertentu untuk rekreasi atau mempelajari keunikan daya tarik tempat wisata yang dikunjungi [4]. Salah satu tempat wisata di Indonesia yang banyak dikunjungi adalah Jawa dengan total kunjungan mencapai 186.894.200 jiwa dalam kurun waktu 5 tahun dimulai dari tahun 2014 hingga 2018. Dalam melakukan sebuah kunjungan wisata biasanya wisatawan akan melakukan proses perencanaan perjalanan dengan melihat informasi yang ada di internet, kemudian wisatawan akan menuliskan tempat tujuan yang akan dikunjungi. Proses merencanakan perjalanan wisata ini dapat membantu dalam mengatur perjalanan agar dapat mengunjungi setiap tempat wisata yang tersedia.

Berdasarkan penelitian sebelumnya membahas mengenai perencanaan perjalanan tempat wisata dan restoran di negara-negara Asia Tenggara[5], penelitian ini menggunakan algoritma *K-Means Clustering* dalam mengatur perencanaan perjalanan. Sistem akan menetapkan tempat yang berdekatan pada hari kunjungan yang sama untuk membantu wisatawan mengurangi waktu dan biaya transportasi yang tidak perlu. Hal ini dilakukan dengan menggunakan algoritma *K-Means Clustering*. Algoritma *K-Means* digunakan untuk mengelompokkan titik-titik yang mirip satu sama lain ke dalam sejumlah *cluster*, k ditentukan oleh penyelenggara perjalanan. Dalam kasus sistem yang diusulkan, k adalah jumlah hari perjalanan. Algoritma ini akan memastikan tempat-tempat yang berdekatan diatur untuk nantinya di kunjungi pada hari yang sama sehingga wisatawan dapat menghemat waktu dan biaya transportasi.

Untuk melihat seberapa optimal algoritma PAM dalam proses *geo-location clustering*, maka penelitian ini akan menggunakan data tempat wisata di Pulau Jawa seperti *longitude*, *latitude*, rating, mushola, toilet, tempat makan, tempat parkir, pemandu wisata, spot foto, pusat informasi wisata, gazebo, kebersihan dan harga tiket. Data tersebut akan digunakan untuk menentukan tempat tujuan wisata. Nilai K yang dimasukkan akan di inisialisasikan sebagai jumlah hari kunjungan. Penelitian ini akan menghasilkan beberapa *cluster* sesuai dengan jumlah hari kunjungan. Algoritma PAM digunakan karena dalam penelitian [5] menyebutkan bahwa terdapat kekurangan pada penelitian tersebut yaitu algoritma *K-Means* yang sensitif terhadap *outlier* sehingga menyebabkan *centroid* yang terbentuk tidak tepat.

II. LANDASAN TEORI

A. *Geo-Location Clustering*

Geo-location adalah identifikasi lokasi geografis dunia nyata dari suatu objek. *Geo-location* mungkin merujuk pada praktik menilai lokasi, atau ke lokasi yang aktual[6]. *Geo-location* dapat digunakan untuk mencari letak geografis suatu objek. Secara garis besar *geo-location clustering* merupakan proses pengelompokan data-data lokasi yang menghasilkan sebuah *cluster* lokasi[7].

B. *Metode Elbow*

Metode *Elbow* merupakan metode yang dinformasi dalam menentukan jumlah *cluster* terbaik dengan cara melihat persentase perbandingan antara jumlah cluster yang akan membentuk siku pada sebuah titik. Metode ini memberikan ide/gagasan dengan cara memilih nilai cluster kemudian menjumlahkan nilai *cluster* untuk dijadikan model data dalam menentukan cluster terbaik. Dan selain itu persentase perhitungan yang dihasilkan merupakan perbandingan antara jumlah cluster yang ditambahkan[8].

$$SSE = \sum_{K=1}^K \sum_{x_i \in C_k} \|x_i - c_k\|_2^2 \quad (1)$$

Keterangan :

SSE : *Sum Of Square Error*

X_i : nilai atribut dari data ke 1

C_k : nilai atribut titik pusat *cluster* ke 1

C. Algoritma PAM Clustering

Algoritma *Partitioning Around Medoids* (PAM) adalah realisasi populer dari pengelompokan *K-Medoids*, menangani masalah dengan cara yang berulang-ulang[8]. Algoritma PAM (*Partitioning Around Medoids*) digunakan dalam menemukan *medoids* sebuah *cluster*. Pemilihan *Medoids* dilakukan secara acak dari objek data k untuk membentuk *cluster* k dan objek data lainnya yang tersisa ditempatkan di dekat *Medoids* dalam sebuah *cluster*[9]. Algoritma PAM mirip dengan *K-Means*, namun algoritma PAM memiliki keunggulan dalam mengatasi kelemahan pada algoritma *K-Means* yang *sensitive* terhadap *noise* dan *outlier*. Keuntungan lainnya adalah hasil dari proses *clustering* tidak bergantung pada urutan *input*[6]. Berikut ini tahapan-tahapan dalam proses *clustering* menggunakan algoritma PAM:

1. Menentukan jumlah cluster atau nilai K.
2. Memilih secara acak *medoids* awal pada sekumpulan data sesuai dengan banyaknya jumlah *cluster*.
3. Hitung jarak untuk setiap data dengan *medoids* awal sehingga menghasilkan jarak terdekat.

$$d_{Euclidean}(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (2)$$

Dimana :

d_{xy} : Jarak data x dan y

n : Jumlah data

x_i : nilai data x ke i

y_i : nilai data y ke i

4. Tandai jarak terdekat data dengan *medoids* awal serta hitung total kedekatan awal.
5. Memilih *medoids* baru pada sekumpulan data sesuai banyaknya jumlah *cluster*.
6. Hitung jarak untuk setiap data dengan *medoids* baru sehingga menghasilkan jarak terdekat.
7. Tandai jarak terdekat data dengan *medoids* baru serta hitung total kedekatan baru.
8. Hitung total simpangan (s) dengan rumus, total kedekatan baru – total kedekatan awal.
9. Jika total simpangan (s) < 0 maka lakukan kembali tahap 5 sampai tahap 8, namun jika total simpangan (s) > 0 maka perulangan dihentikan sehingga mendapatkan *cluster* dengan anggota *cluster* tetap[10].

D. Silhouette Coefficient

Tahap evaluasi merupakan tahap terakhir yang dilakukan, pada tahap ini evaluasi *Silhouette Coefficient* dilakukan. Evaluasi ini menggabungkan 2 metode, yaitu metode *Cohesion* dan metode *Separation*. Metode *Cohesion* digunakan untuk mengukur jarak antara objek satu dengan objek lainnya didalam sebuah *cluster*, sedangkan metode *Separation* digunakan untuk mengukur jarak antara *cluster* dengan *cluster* yang lain[11]. Ada 3 tahapan dalam perhitungan *Silhouette Coefficient* yaitu diantaranya[12]:

1. Untuk setiap objek i, skor rata-rata titik satu dihitung dengan semua objek yang ada dalam *cluster* tersebut. Maka nilai rata-rata akan didapatkan atau ditunjukkan sebagai a_i .
2. Untuk objek i, dihitung jarak rata-rata minimum dari satu titik ke titik yang lainnya dengan *cluster* yang berbeda. Sehingga akan didapatkan rata-rata minimum atau ditunjukkan sebagai b_i .
3. Setelah semua nilai diketahui, nilai *Silhouette Coefficient* dapat ditentukan dengan menggunakan persamaan:

$$S_i = \frac{b_i - a_i}{\max(a_i - b_i)} \quad (3)$$

Keterangan :

b : Jarak rata-rata *medoid* dengan obyek diluar *cluster*

a : Jarak rata-rata antara *medoid* dengan objek didalam *cluster*

Adapun nilai *Silhouette Coefficient* pada angka antara -1 sampai dengan 1. Berikut merupakan *Silhouette Coefficient* menurut Kaufman dan Rousseeuw:

$0.7 < SC \leq 1$ merupakan *cluster Strong Structure*

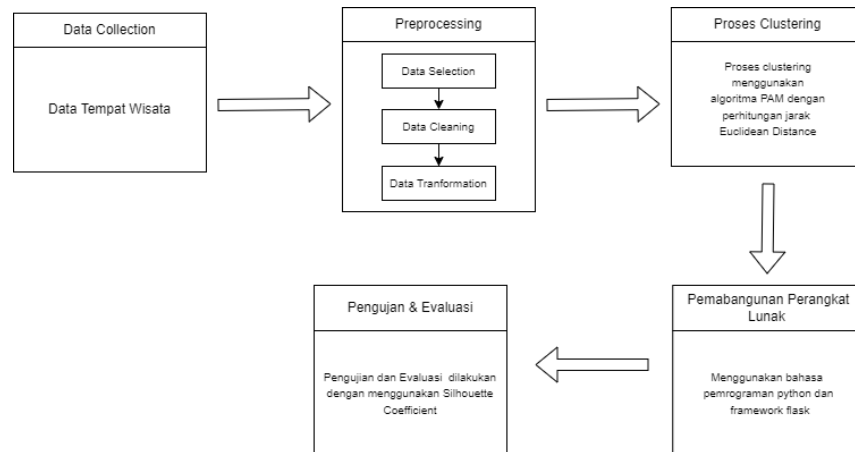
$0.5 < SC \leq 0.7$ merupakan *cluster Medium Structure*

$0.25 < SC \leq 0.5$ merupakan *cluster Weak Structure*

$SC \leq 0.25$ merupakan *cluster No Structure*

III. METODELOGI PENELITIAN

Pada penelitian ini dilakukan tahapan-tahapan untuk mendapatkan hasil yang diinginkan. Berikut tahapan penelitian yang dapat dilihat pada gambar 1.



Gambar 1 Metode Penelitian

A. Data Collection

Pengumpulan data yang dilakukan pada penelitian ini yaitu dengan melakukan proses *input* data manual yang didapatkan dari Gmaps dan idntrip.

B. Data Pre-Processing

Proses data *pre-processing* dilakukan dengan 3 tahapan yaitu tahapan data *cleaning*, data *selection*, dan data *transformation*[13]. Data *selections* merupakan sekumpulan data yang diseleksi untuk dapat menciptakan kelompok data target[14]. Data *cleaning* merupakan tahapan untuk menghindari terjadinya redundan pada adanya *missing value*. Data *transformation* merupakan sebuah proses integrasi berdasarkan data yang sudah dipilih sehingga data yang dimiliki sesuai dengan proses *data mining* yang akan dilakukan selanjutnya[15].

C. Proses Clustering

Pada tahapan sebelumnya menghasilkan data yang siap diuji, sehingga pada tahapan ini data tersebut akan dilakukan proses data *mining* dengan menggunakan metode *clustering* berupa PAM (*Partitioning Around Medoids*). Dimana dalam metode ini data yang memiliki karakteristik yang sama akan dikelompokkan dalam satu kelompok dan data yang memiliki karakteristik berbeda akan dikelompokkan dengan kelompok lain yang sesuai dengan karakteristik data tersebut.

D. Pembangunan Perangkat Lunak

Pada tahapan ini, dilakukan proses pembangunan perangkat lunak. Tahapan ini dilakukan untuk mempermudah dalam proses *clustering* dan visualisasi data yang digunakan. Pembangunan perangkat lunak dibangun dengan menggunakan bahasa pemrograman *python* dan *framework* yang digunakan yaitu *flask*.

E. Pengujian dan Evaluasi

Pada tahap ini merupakan tahap mengidentifikasi hasil *cluster* yang akan terbentuk dengan mengevaluasi nilai *Silhouette Score* menggunakan pengujian *Silhouette Coefficient*.

IV. PENGUJIAN DAN ANALISIS

Pada penelitian ini data yang digunakan sebanyak 1014 data dengan total 13 atribut. Adapun atribut yang digunakan terdiri dari *longitude*, *latitude*, rating, toilet, tempat makan, mushola, tempat parkir, pemandu wisata, spot foto, pusat informasi wisata, harga tiket, kebersihan dan gazebo.

Tabel 1 Data Tempat Wisata

Rating	Mushola	Toilet	Tempat Makan	Tempat Parkir	Harga Tiket	Latitude (y)
4.5	Ya	Ya	Ya	Ya	185.000	-7.4164463
4.5	Ya	Ya	Ya	Ya	15.000	-7.3650621
4.1	Ya	Ya	Ya	Ya	20.000	-7.1940697
4.4	Ya	Ya	Ya	Ya	5.000	-7.1894444
4.4	Tidak	Ya	Ya	Ya	20.000	-7.2201011
4.3	Ya	Ya	Ya	Ya	25.000	-7.389256
4.5	Tidak	Ya	Ya	Ya	185.000	-7.3889466
4.4	Ya	Ya	Ya	Ya	185.000	-7.390375
4.1	Ya	Ya	Ya	Ya	120.000	-7.3905592
4.1	Tidak	Ya	Tidak	Ya	5.000	-7.1972222
4.6	Ya	Ya	Tidak	Ya	25.000	-7.3885367
4.3	Ya	Ya	Tidak	Ya	4.000	-7.4142818
....
4.3	Ya	Ya	Ya	Ya	15.000	-7.3586273

A. Data Pre-processing

Pada bagian data *pre-processing* didalamnya terdapat 3 tahapan, yaitu data *selection*, data *cleaning*, dan data *transformation*. Dalam penelitian ini, semua atribut pada data digunakan meliputi *longitude*, *latitude*, rating, tempat makan, mushola, toilet, tempat parkir, pemandu wisata, spot foto, pusat informasi wisata, gazebo, harga tiket, dan kebersihan. Untuk proses data *cleaning* dilakukan dengan menggunakan bahasa pemrograman *python* pada *google colabs*. Dari hasil data *cleaning* yang dilakukan, tidak ada *missing value* pada data sehingga tidak dilakukan proses *handling missing value*.

Pada tahapan data *transformation* dilakukan proses transformasi data, karena penelitian ini menggunakan metode *clustering* sehingga data dalam bentuk kategorik perlu diubah menjadi data dalam bentuk numerik. Ada beberapa atribut yang masuk kedalam data kategorik diantaranya Mushola, Toilet, Tempat makan, Tempat parkir, Pemandu wisata, Spot foto, Pusat informasi wisata, Gazebo dan Kebersihan. Pada atribut Kebersihan yang memiliki nilai Baik dan Kurang diubah menjadi Baik bernilai 1 dan Kurang bernilai 0, sedangkan untuk data yang bernilai Ya dan Tidak diubah menjadi, Ya = 1 dan Tidak = 0. Setelah semua data berbentuk numerik, selanjutnya dilakukan proses normalisasi data agar nilai pada setiap data memiliki rentang yang sama atau tidak ada nilai yang terlalu besar maupun nilai yang terlalu kecil, sehingga data berada pada *range* 0 sampai 1.

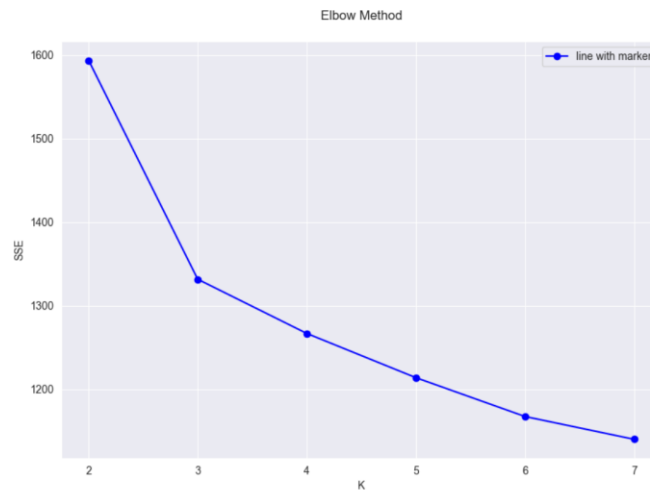
Tabel 2 Data Siap Uji

Rating	Mushola	Toilet	Tempat Makan	Tempat Parkir	Harga Tiket	Latitude (y)
0.7500	1.0000	1.0000	1.0000	1.0000	0.8222	109.8172
0.7500	1.0000	1.0000	1.0000	1.0000	0.0667	109.8951
0.5500	1.0000	1.0000	1.0000	1.0000	0.0889	109.8807
0.7000	1.0000	1.0000	1.0000	1.0000	0.0222	109.8522

0.7000	0.0000	1.0000	1.0000	1.0000	0.0889	109.9050
0.6500	1.0000	1.0000	1.0000	1.0000	0.1111	109.6908
0.7500	0.0000	1.0000	1.0000	1.0000	0.8222	109.7441
0.7000	1.0000	1.0000	1.0000	1.0000	0.8222	109.7448
0.5500	1.0000	1.0000	1.0000	1.0000	0.5333	109.6869
0.5500	0.0000	1.0000	0.0000	1.0000	0.0222	109.8519
0.8000	1.0000	1.0000	0.0000	1.0000	0.1111	109.6907
0.6500	1.0000	1.0000	0.0000	1.0000	0.0178	109.7573
....
0.6500	1.0000	1.0000	1.0000	1.0000	0.0667	109.2426

B. Penentuan Cluster

Metode *Elbow* digunakan untuk menentukan jumlah K atau nilai *cluster* yang optimal dari data tempat wisata. Metode ini dilakukan dengan cara melihat hasil *Sum of Square Error* (SSE) dari setiap *cluster*, apabila *cluster* memiliki jumlah penurunan SSE yang signifikan maka dapat di tentukan jumlah *cluster* yang optimalnya. Berikut tahap-tahap yang dilakukan untuk menghitung nilai SSE dengan rumus seperti pada persamaan (1).

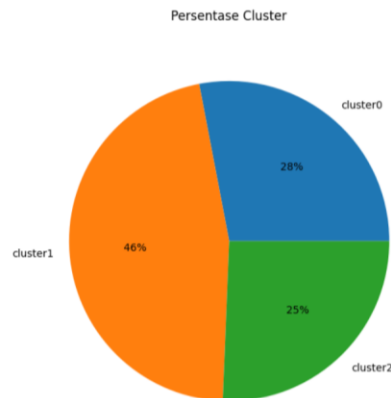


Gambar 2 Grafik Metode Elbow

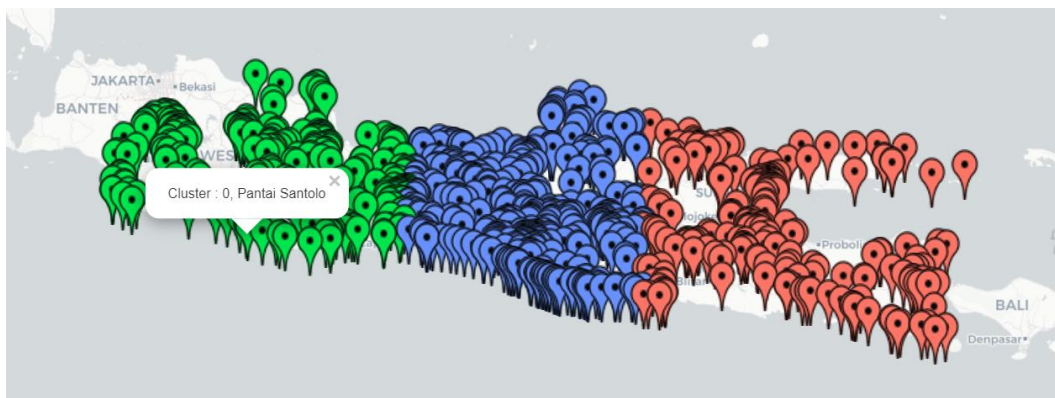
Pada gambar 2 yang merupakan grafik dari hasil perhitungan Metode *Elbow*, didapatkan bahwa nilai K atau nilai *cluster* yang optimal yaitu 3. Pada titik 3 memiliki sudut terkecil diantara titik yang lainnya, sehingga nilai K yang digunakan yaitu K=3.

C. Hasil Clustering

Clustering yang dilakukan untuk mengelompokan data tempat wisata ini menggunakan algoritma PAM. Setelah mendapatkan nilai K yang optimal yaitu K=3, berikut adalah hasil *presentase* clustering yang terbentuk ditunjukkan pada Gambar 3 dan Gambar 4 merupakan hasil visualisasi dalam bentuk peta.



Gambar 3 Persentase Hasil Cluster



Gambar 4 Visualisasi Cluster dengan Peta

D. Pengujian *Clustering*

Pada penelitian ini uji evaluasi hasil *cluster* menggunakan perhitungan *Silhouette Coefficient*. Hasil *cluster* terbaik dilihat berdasarkan jarak antar objek satu dengan objek lainnya didalam sebuah *cluster* dan berdasarkan jarak antar *cluster* dengan *cluster* lainnya. Hasil nilai *silhouette coefficient* ditunjukkan pada Gambar 5.

Cluster	Silhouette Coefficient
3	0.33863890606304814

Gambar 5 Silhouette Coefficient

Berdasarkan hasil perhitungan yang telah dilakukan, didapatkan bahwa nilai $K=3$ memiliki nilai *silhouette* yaitu 0.3386. Dimana dengan nilai yang didapat tersebut tergolong kedalam *Cluster Weak Structure*.

E. Interpretasi

Dari hasil *cluster* yang terbentuk yaitu $K=3$, berikut analisa interpretasi untuk setiap *cluster*:

- *Cluster 0*, pada *cluster* ini memiliki jumlah total kunjungan sebanyak 284 tempat wisata yang memiliki karakteristik nilai dari atribut rating dengan nilai rata-rata 4.33, atribut kebersihan dengan nilai rata-rata 0.75, serta atribut harga tiket dengan harga maksimal yaitu Rp 170.000,00. Selain itu pada *cluster* ini rata-rata memiliki fasilitas mushola, toilet, tempat makan, tempat parkir dan spot foto. Untuk rata-rata koordinat *cluster 0* masuk kedalam wilayah Jawa Barat.

- *Cluster 1*, pada *cluster* ini memiliki jumlah total kunjungan sebanyak 470 tempat wisata yang memiliki karakteristik nilai dari atribut rating yaitu dengan nilai rata-rata 4.34, atribut kebersihan dengan nilai rata-rata 0.74, serta atribut harga tiket dengan harga maksimal Rp 200.000,00. Selain itu pada *cluster* ini *cluster* ini rata-rata memiliki fasilitas mushola, toilet, tempat makan, tempat parkir, pemandu wisata dan spot foto. Untuk rata-rata koordinat *cluster 1* masuk kedalam wilayah Jawa Tengah.
- *Cluster 2*, pada *cluster* ini memiliki jumlah total kunjungan sebanyak 260 tempat wisata yang memiliki karakteristik nilai dari atribut rating yaitu dengan nilai rata-rata 4.35, atribut kebersihan dengan nilai rata-rata 0.83, serta atribut harga tiket dengan harga maksimal Rp 225.000,00. Selain itu pada *cluster* ini rata-rata memiliki fasilitas mushola, toilet, tempat makan, tempat parkir, pemandu wisata dan spot foto. Untuk rata-rata koordinat *cluster 2* masuk kedalam wilayah Jawa Timur.

V. KESIMPULAN

Dalam penelitian *Geo-Location Clustering* menggunakan Algoritma PAM untuk menentukan tempat tujuan wisata dapat menghasilkan beberapa *cluster*. Pada penelitian ini data yang digunakan berjumlah 1014 record dan 13 variable, untuk menentukan nilai K yang optimal digunakan metode *elbow*, dari metode *elbow* tersebut didapatkan nilai K yang optimal yaitu 3. Hasil dari *cluster* tersebut memiliki karakteristik atribut untuk setiap *cluster*-nya. *Cluster 0* memiliki jumlah kunjungan wisata sebanyak 284 tempat wisata dengan rata-rata rating yaitu 4.33 dan rata-rata koordinat masuk kedalam wilayah Jawa Barat, *cluster 1* memiliki jumlah kunjungan wisata sebanyak 470 tempat wisata dengan rata-rata rating yaitu 4.34 dan rata-rata koordinat masuk kedalam wilayah Jawa Tengah, dan *cluster 2* memiliki jumlah kunjungan wisata sebanyak 260 tempat wisata dengan rata-rata rating yaitu 4.35 dan rata-rata koordinat masuk kedalam wilayah Jawa Timur. Hasil evaluasi untuk penelitian ini menggunakan metode *Silhouette Coefficient* yang menghasilkan nilai *cluster* terbaik yaitu 3 dengan nilai *silhouette* sebesar 0.3386, sehingga *cluster* yang terbentuk cukup baik. *Cluster* yang terbentuk akan bertambah jika data yang dimiliki juga bertambah. Dari hasil evaluasi tersebut dapat ditarik kesimpulan bahwa algoritma PAM cukup baik untuk diterapkan kedalam proses *geo-location clustering*.

REFERENSI

- [1] S. Agarwal, "Data mining: Data mining concepts and techniques," *Proceedings - 2013 International Conference on Machine Intelligence Research and Advancement, ICMIRA 2013*, pp. 203–207, 2014, doi: 10.1109/ICMIRA.2013.45.
- [2] D. Marlina, N. Lina, A. Fernando, and A. Ramadhan, "Implementasi Algoritma K-Medoids dan K-Means untuk Pengelompokan Wilayah Sebaran Cacat pada Anak," *Jurnal CoreIT: Jurnal Hasil Penelitian Ilmu Komputer dan Teknologi Informasi*, vol. 4, no. 2, p. 64, 2018, doi: 10.24014/coreit.v4i2.4498.
- [3] F. Lin, J. Jiang, J. Fan, and S. Wang, "A stacking model for variation prediction of public bicycle traffic flow," *Intelligent Data Analysis*, vol. 22, no. 4, pp. 911–933, 2018, doi: 10.3233/IDA-173443.
- [4] N. Yanishevskaya, L. Kuznetsova, A. Zhigalov, D. Parfenov, and I. Bolodurina, "Development of an intellectual module for selection of places to travel in the virtual assistant system for planning trips," *Journal of Physics: Conference Series*, vol. 1399, no. 3, 2019, doi: 10.1088/1742-6596/1399/3/033059.
- [5] L. Yen Yee, S. M. Letchumy Belaidan, N. Azlina Abd Rahman, and K. Shajaratuddin Harun, "Implementing K-Means Clustering Algorithm in Collaborative Trip Advisory and Planning System," vol. 7, no. 2, pp. 723–740, 2019, [Online]. Available: <http://pen.ius.edu.ba>
- [6] W. Zhang and S. V. Ukkusuri, "Share-a-Cab: Scalable clustering taxi group ride stand from huge geolocation data," *IEEE Access*, vol. 9, pp. 9771–9776, 2021, doi: 10.1109/ACCESS.2021.3050299.
- [7] Jakimavičius, Palevičius, Antuchevičienė, and Karpavičius, "Internet GIS-Based Multimodal Public Transport Trip Planning Information System for Travelers in Lithuania," *ISPRS International Journal of Geo-Information*, vol. 8, no. 8, p. 319, 2019, doi: 10.3390/ijgi8080319.
- [8] R. Nainggolan, R. Perangin-Angin, E. Simarmata, and A. F. Tarigan, "Improved the Performance of the K-Means Cluster Using the Sum of Squared Error (SSE) optimized by using the Elbow Method," *Journal of Physics: Conference Series*, vol. 1361, no. 1, 2019, doi: 10.1088/1742-6596/1361/1/012015.
- [9] D. F. Pramesti, Lahan, M. Tanzil Furqon, and C. Dewi, "Implementasi Metode K-Medoids Clustering Untuk Pengelompokan Data Potensi Kebakaran Hutan/Lahan Berdasarkan Persebaran Titik Panas (Hotspot)," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 1, no. 9, pp. 723–732, 2017, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [10] P. Arora, Deepali, and S. Varshney, "Analysis of K-Means and K-Medoids Algorithm for Big Data," *Physics Procedia*, vol. 78, no. December 2015, pp. 507–512, 2016, doi: 10.1016/j.procs.2016.02.095.
- [11] H. Řezanková, "Different approaches to the silhouette coefficient calculation in cluster evaluation," *21st International Scientific Conference AMSE*, no. September, pp. 1–10, 2018.
- [12] M. Pant, T. Radha, and V. P. Singh, "Particle swarm optimization using Gaussian inertia weight," *Proceedings - International Conference on Computational Intelligence and Multimedia Applications, ICCIMA 2007*, vol. 1, pp. 97–102, 2008, doi: 10.1109/ICCIMA.2007.328.
- [13] R. Raksanagara, Y. H. Chrisnanto, and A. I. Hadiana, "Analisis Sentimen Jasa Ekspedisi Barang Menggunakan Metode Naïve Bayes," *Prosiding SNST ke-7 Tahun 2016*, pp. 19–24, 2016.
- [14] H. M. Safhi, B. Frikh, and B. Ouhbi, "Assessing reliability of Big Data Knowledge Discovery process," *Procedia Computer Science*, vol. 148, pp. 30–36, 2019, doi: 10.1016/j.procs.2019.01.005.
- [15] Z. Ge, Z. Song, S. X. Ding, and B. Huang, "Data Mining and Analytics in the Process Industry: The Role of Machine Learning," *IEEE Access*, vol. 5, pp. 20590–20616, 2017, doi: 10.1109/ACCESS.2017.2756872.