

Evaluasi Kinerja Algoritma K-Means dan K-Medoids untuk Klasterisasi Destinasi Wisata Bali

Aryanti^{1*}, Dian Anjani², Muhammad Gian Azzra Ramadhan³, Suci Putri Widyani⁴,
Adinda Kamila Sari⁵

^{1,2,3,4,5}Program Studi Sarjana Terapan Teknik Telekomunikasi, Politeknik Negeri Sriwijaya

*aryanti@polsri.ac.id

Abstrak

Penelitian ini bertujuan untuk membandingkan kinerja algoritma K-Means dan K-Medoids dalam mengelompokkan destinasi wisata di Provinsi Bali berdasarkan karakteristik multidimensi, meliputi rating, jumlah ulasan, latitude, longitude, kategori, dan kabupaten/kota. Dataset yang digunakan berasal dari Kaggle sebanyak 1.000 data destinasi wisata yang telah melalui proses preprocessing berupa pembersihan data, seleksi fitur, normalisasi, dan transformasi logaritmik. Jumlah cluster optimal ditentukan menggunakan Metode Elbow pada rentang K=2 hingga K=10, sedangkan kualitas cluster dievaluasi menggunakan Silhouette Score, Davies-Bouldin Index (DBI), dan Mean Squared Error (MSE). Hasil penelitian menunjukkan bahwa jumlah cluster optimal adalah K=10. K-Means menghasilkan Silhouette Score sebesar 0,2078, DBI sebesar 1,2714, dan MSE sebesar 2,2764, sehingga menunjukkan kinerja yang lebih baik dibandingkan K-Medoids. Penelitian ini memberikan kontribusi berupa evaluasi komparatif kedua algoritma serta rekomendasi metode clustering yang lebih sesuai untuk data destinasi wisata Bali. Hasil penelitian juga dapat mendukung pengambilan keputusan dalam pengelolaan dan pengembangan sektor pariwisata.

Kata kunci : Clustering, K-Means, K-Medoids, Elbow, Silhouette Score, DBI, Pariwisata.

Abstract

This study aims to compare the performance of the K-Means and K-Medoids algorithms in clustering tourist destinations in Bali Province based on multidimensional characteristics, including rating, number of reviews, latitude, longitude, category, and district/city. The dataset used comes from Kaggle with 1,000 tourist destination data that have gone through a preprocessing process in the form of data cleaning, feature selection, normalization, and logarithmic transformation. The optimal number of clusters is determined using the Elbow Method in the range of K = 2 to K = 10, while the quality of the clusters is evaluated using the Silhouette Score, Davies-Bouldin Index (DBI), and Mean Squared Error (MSE). The results show that the optimal number of clusters is K = 10. K-Means produces a Silhouette Score of 0.2078, a DBI of 1.2714, and an MSE of 2.2764, thus showing better performance than K-Medoids. This study contributes in the form of a comparative evaluation of the two algorithms and recommendations for clustering methods that are more suitable for Bali tourist destination data. The research results can also support decision making in the management and development of the tourism sector.

Keywords : Clustering, K-Means, K-Medoids, Elbow, Silhouette Score, DBI, Tourism.

1. Pendahuluan

Provinsi Bali merupakan salah satu destinasi wisata unggulan di Indonesia yang memiliki kontribusi besar terhadap perekonomian nasional. Tingginya jumlah kunjungan wisatawan menjadikan Bali sebagai pusat aktivitas

pariwisata yang strategis, sekaligus menghasilkan data yang terus berkembang dan beragam^[1]. Hal ini juga menunjukkan bahwa data yang besar dan beragam memerlukan teknik analisis yang tepat, seperti clustering untuk

menyederhanakan data dan menghasilkan informasi yang lebih bermakna [2].

Namun, pengelolaan destinasi wisata di Bali menghadapi tantangan dalam memahami pola kunjungan wisatawan yang kompleks. Perbedaan tingkat popularitas antar destinasi serta perubahan tren wisata menyebabkan pengelompokan destinasi menjadi tidak mudah. Tanpa analisis yang tepat, pengambilan keputusan terkait pengembangan dan pemasaran pariwisata menjadi kurang optimal[3]. Pendekatan clustering sering dimanfaatkan sebagai media untuk menemukan atau menganalisis pola tersembunyi dalam data, seperti pengelompokan berdasarkan karakteristik tertentu. K-Means merupakan algoritma clustering yang banyak diterapkan, di mana data dikelompokkan menjadi sejumlah cluster berdasarkan karakteristiknya[4].

K-Means menjadi algoritma yang diterapkan karena prosesnya tidak kompleks serta efektif dalam mengolah data berukuran besar[1]. Namun, metode ini memiliki kelemahan, seperti sensitif terhadap inisialisasi centroid dan rentan terhadap outlier[5]. Sebagai alternatif, algoritma K-Medoids menawarkan keunggulan dalam ketahanan terhadap outlier karena menggunakan data aktual sebagai pusat cluster[6]. K-Means maupun K-Medoids memiliki karakteristik berbeda, di mana K-Means unggul dalam

efisiensi komputasi, sedangkan K-Medoids lebih stabil terhadap outlier[7]. [8].

Meskipun berbagai penelitian telah menerapkan metode clustering pada sektor pariwisata maupun membandingkan algoritma clustering pada bidang lain, penelitian yang secara khusus membandingkan performa K-Means dan K-Medoids pada data destinasi wisata di Provinsi Bali masih terbatas[6],[7],[8],[9],[10]. Selain itu, sebagian besar penelitian terdahulu berfokus pada penerapan salah satu algoritma tanpa melakukan evaluasi komparatif secara langsung. Kondisi tersebut menunjukkan perlunya penelitian yang dapat memberikan gambaran mengenai metode yang lebih sesuai untuk karakteristik data destinasi wisata Bali.

Berdasarkan uraian tersebut, studi ini bertujuan untuk menganalisis serta membandingkan kinerja algoritma K-Means dan K-Medoids dalam pengelompokan tempat wisata di Provinsi Bali.. Hasil yang didapat diharapkan dapat menjadi acuan dalam memilih metode clustering yang tepat serta mendukung proses pengambilan keputusan di sektor pengelolaan pariwisata.

2. Tinjauan Pustaka

2.1. Penelitian Terkait

Sejumlah penelitian terdahulu memiliki keterkaitan dan digunakan sebagai dasar pembandingan.

1. Studi sebelumnya menggunakan teknik K-Means untuk mengklasifikasikan data kunjungan wisata Bali berdasarkan ulasan Tripadvisor, berhasil mengelompokkan destinasi menjadi kategori tinggi, sedang, dan rendah. Penelitian ini relevan karena menggunakan metode clustering pada data pariwisata^[1].

2. Penelitian lain menggunakan metode clustering untuk menganalisis persebaran hotel di Bali berbasis lokasi geografis dan membuktikan clustering mampu mengidentifikasi pola distribusi wilayah wisata. Penelitian ini relevan karena sama-sama menggunakan pendekatan clustering dalam sektor pariwisata^[3].

3. Penelitian berikutnya melakukan analisis berbagai algoritma clustering menunjukkan bahwa kinerja algoritma bervariasi tergantung pada dataset yang digunakan. Temuan penelitian membuktikan masing-masing algoritma memiliki keunggulan dan keterbatasan yang dipengaruhi oleh sifat data, sehingga menekankan pentingnya pemilihan metode yang tepat^[11].

4. Studi lain membahas secara komprehensif tentang K-Means menunjukkan bahwa algoritma ini memiliki kelebihan dan keterbatasan dalam proses clustering. Keterbatasan utama K-Means adalah rentan terhadap inisialisasi awal dan keberadaan outlier, yang menjadikan penelitian tersebut penting sebagai dasar dalam memahami penggunaan K-Means^[5].

5. Studi tambahan membandingkan kedua algoritma untuk proses clustering. Temuan menunjukkan bahwa K-Medoids lebih efisien untuk komputasi serta lebih stabil terhadap outlier. Penelitian ini relevan karena membandingkan kedua metode penelitian yang digunakan^[6]

2.2. Landasan Teori

1. Clustering.

Dalam penelitian ini, metode pengelompokan data berdasarkan tingkat kemiripan karakteristik digunakan untuk mengelompokkan destinasi wisata di Provinsi Bali. Pola data dapat dianalisis dan dibandingkan antar algoritma dengan menggunakan clustering data mining^[11].

2. Data Mining

Data mining digunakan dalam penelitian ini untuk mengolah dataset destinasi wisata dan menghasilkan informasi penting, ini dilakukan dengan menggunakan algoritma pengelompokan untuk mendapatkan pola pengelompokan data^[12].

3. K-Medoids

K-Medoids merupakan algoritma clustering yang menggunakan data aktual sebagai pusat cluster (medoid). Metode ini meminimalkan jarak total antara medoid dan setiap data dalam satu cluster, yang membuatnya lebih stabil terhadap pengaruh outlier daripada metode berbasis rata-rata^{[13], [9]}:

$$\sum_{i=1}^k \sum_{x \in C_i} d(x, m_i) \quad (1)$$

Keterangan:

k : jumlah cluster

C_i : himpunan data cluster ke-i

x : data dalam cluster

m_i : medoid (titik pusat cluster)

$d(x, m_i)$: jarak antara data dengan medoid, umumnya menggunakan Euclidean distance

4. K-Means

Pada K-Means, pengelompokan dilakukan berdasarkan jarak data terhadap centroid yang mencerminkan rata-rata anggota cluster. Proses ini bertujuan untuk mengurangi penyebaran data dalam cluster agar terbentuk kelompok yang lebih seragam^{[5], [10]}.

$$\sum_{i=1}^k \sum_{x \in C_i} \|x - \mu_i\|^2 \quad (2)$$

Keterangan:

k : jumlah cluster

C_i : himpunan data cluster ke-i

x : data dalam cluster

μ_i : centroid atau rata-rata dari cluster ke-i

$\|x - \mu_i\|^2$ = jarak kuadrat antara data dan centroid (biasanya menggunakan Euclidean distance)

5. Davies-Bouldin Index

Davies-Bouldin Index (DBI) digunakan sebagai untuk membandingkan jarak dalam cluster dan jarak antar cluster ^[14].

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{j \neq i} \left(\frac{S_i + S_j}{M_{ij}} \right) \quad (3)$$

Keterangan:

k : jumlah cluster

S_i : rata-rata jarak data ke centroid pada cluster ke-i

S_j : himpunan data pada cluster ke-j

M_{ij} : jarak antar centroid cluster ke-i dan ke-j

6. Silhouette Score

Silhouette Score digunakan dalam penelitian ini untuk mengevaluasi kualitas pengelompokan. Nilai yang rendah menunjukkan kemungkinan kesalahan pengelompokan, sedangkan nilai yang mendekati satu menunjukkan bahwa data berada pada cluster yang tepat^[14].

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (4)$$

Keterangan:

$s(i)$: nilai silhouette untuk data ke-i

$a(i)$: rata-rata jarak data ke cluster yang sama $b(i)$

: rata-rata jarak ke cluster terdekat

7. Metode Elbow

Metode Elbow menentukan jumlah cluster yang ideal untuk digunakan. Apabila penurunan nilai mulai melambat menunjukkan jumlah cluster yang paling sesuai^[15].

$$SSE(k) = \sum_{j=1}^k \sum_{X_i \in S_j} \|X_i - C_{S_j}\|^2 \quad (5)$$

Keterangan:

k : jumlah cluster

S_j : himpunan data pada cluster ke- j

X_i : data ke-1

C_j : centroid (pusat cluster je-j)

$||X_i - C_j||^2$: jarak kuadrat antara data dan centroid

3. Metode Penelitian

3.1 Pendekatan Penelitian

Metode kuantitatif diterapkan untuk menilai efektivitas masing-masing algoritma dalam pengelompokan data destinasi wisata di Provinsi Bali. Hasil pengelompokan dari dua algoritma dianalisis menggunakan metrik validitas objektif untuk mencari algoritma yang paling efisien dalam menghasilkan rekomendasi peta area wisata yang lebih tepat.

3.2 Prosedur Penelitian

Penelitian ini dilakukan melalui enam tahapan berurutan, dimulai dari pengumpulan data destinasi wisata yang dilanjutkan dengan prapemrosesan untuk meningkatkan kualitas data. Jumlah klaster ditentukan menggunakan metode Elbow sebelum proses pengelompokan dengan algoritma yang dibandingkan. Hasil clustering kemudian dievaluasi menggunakan beberapa metrik validitas, dan tahap akhir dilakukan analisis komparatif untuk menilai kinerja kedua algoritma dalam pemetaan destinasi wisata di Provinsi Bali.

3.3 Sumber Dataset

Dataset yang dipakai berasal dari “Bali Tourist Destinations” di platform Kaggle dan terdiri dari lebih dari 1.000 entri dengan 11 atribut. Penelitian ini menggunakan 1.000 data lokasi wisata sebagai sampel. Dari semua atribut yang ada, `nama_tempat_wisata` berfungsi untuk mengidentifikasi objek, sementara `rating` dipilih sebagai parameter numerik dalam proses clustering karena mencerminkan penilaian pengunjung terhadap setiap destinasi.

3.4 Pengumpulan Data

Data penelitian diperoleh melalui studi dokumentasi menggunakan dataset destinasi wisata di Bali yang tersedia di platform Kaggle dalam format .csv. Dari total data yang ada, diambil 100 entri sebagai sampel penelitian untuk mendukung analisis perbandingan antara algoritma K-Means dan K-Medoids. Sebelum tahap pengelompokan, atribut `rating` diperiksa untuk memastikan tidak ada nilai yang hilang agar kualitas data tetap terjaga selama analisis.

3.5 Teknik Analisis Data

Merujuk pada diagram di Gambar 2, langkah-langkah analisis data dalam studi ini dapat diuraikan sebagai berikut.

1) Tahap Persiapan dan Preprocessing: Analisis dimulai dengan dataset yang terdiri dari 1.000 destinasi wisata di Provinsi Bali. Data

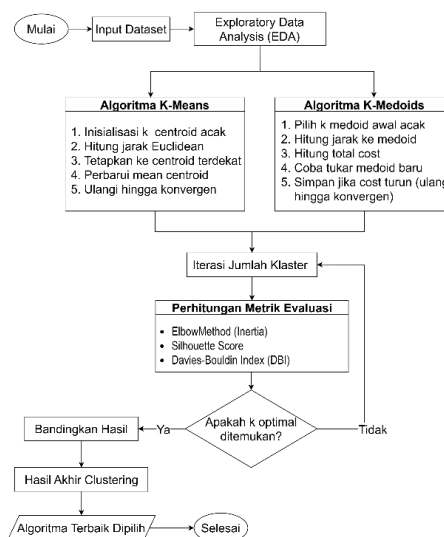
dieksplorasi menggunakan EDA, kemudian atribut jumlah ulasan ditransformasikan dengan fungsi logaritma untuk menstabilkan distribusi.

2) Penerapan Algoritma K-Means dan K-Medoids: Dataset yang telah diproses dianalisis menggunakan kedua algoritma dengan variasi jumlah kluster K=2 hingga K=10 untuk memperoleh konfigurasi optimal.

3) Evaluasi Hasil Clustering: Hasil pengelompokan dievaluasi menggunakan Elbow Method, Silhouette Coefficient, Davies-Bouldin Index (DBI), dan Mean Squared Error (MSE) untuk menilai kualitas pemisahan dan homogenitas kluster.

4) Penentuan Model Terbaik: Perbandingan hasil kedua algoritma dilakukan berdasarkan nilai metrik evaluasi untuk menentukan kinerja terbaik.

Validasi Spasial dan Hasil Akhir: Hasil clustering divisualisasikan dalam bentuk peta menggunakan koordinat lintang dan bujur guna melihat kesesuaian pola kluster dengan kondisi geografis. Analisis dilakukan menggunakan Python dengan pendekatan komparatif eksperimental untuk membandingkan kinerja K-Means dan K-Medoids dalam pengelompokan destinasi wisata di Bali.



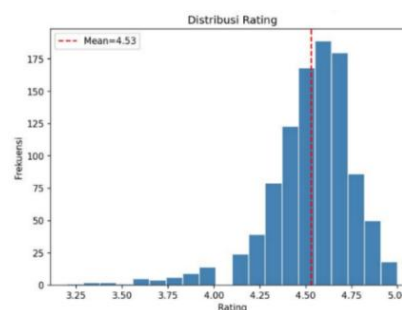
Gambar 2. Flowchart Analisis Data

4. Hasil dan Pembahasan

4.1. Hasil Penelitian

1. Exploratory Data Analysis

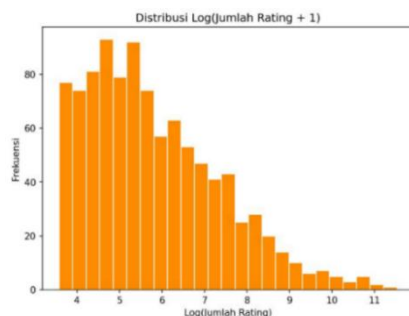
Kumpulan data penelitian mencakup 1.000 lokasi wisata yang tersebar di sembilan daerah administratif dalam Provinsi Bali dan dianalisis dengan empat atribut utama.



Gambar 3. Distribusi Rating Destinasi Wisata

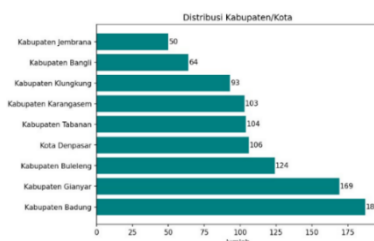
Rata-rata nilai rating yang tercatat adalah 4,53 dari skala maksimum 5,0 dengan pola distribusi yang agak miring ke kiri. Sebagian besar data terkumpul pada rentang 4,50–4,75, menunjukkan bahwa sebagian besar destinasi wisata

mendapatkan penilaian yang baik dari para pengunjung. Fokus data pada kisaran nilai yang sempit mengakibatkan variasi antar objek menjadi terbatas, sehingga pemisahan cluster bisa kurang optimal jika analisis hanya mengandalkan atribut rating.



Gambar 4. Distribusi Jumlah Rating

Sebaran data untuk variabel jumlah ulasan menunjukkan asimetri yang signifikan dengan dominasi beberapa nilai ekstrem. Sehingga, transformasi logaritmik dengan $\text{Log}(\text{Jumlah Rating} + 1)$ digunakan untuk memperbaiki sifat distribusi data. Setelah dilakukan transformasi, pola distribusi menjadi lebih terfokus dan mendekati distribusi log-normal. Konsentrasi data yang paling tinggi terletak di rentang log 4–6, menunjukkan bahwa sebagian besar destinasi pariwisata mendapatkan ulasan dalam jumlah sedang.

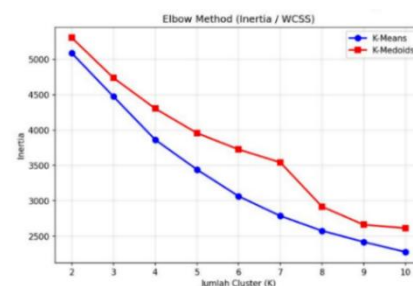


Gambar 5. Distribusi destinasi wisata

Distribusi lokasi wisata antar kabupaten/kota menunjukkan ketidakseimbangan yang sangat terlihat. Badung dan Gianyar merupakan daerah dengan jumlah destinasi terbanyak, yaitu 187 (18,7%) dan 169 (16,9%). Sebaliknya, Jembrana dan Bangli memiliki angka yang lebih sedikit, yaitu 50 (5,0%) dan 64 (6,4%). Hasil penelitian ini mengindikasikan bahwa distribusi tempat wisata di Bali lebih terfokus di bagian selatan dan tengah.

2. Menentukan Jumlah Cluster Optimal (K)

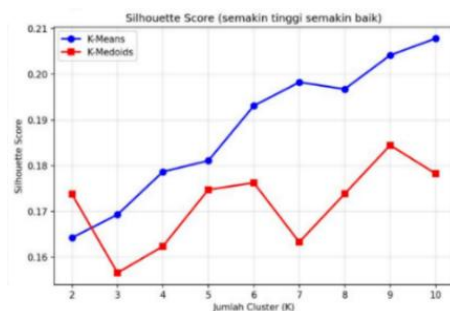
Menentukan jumlah kluster yang optimal dilakukan dengan menganalisis beberapa metrik secara bersamaan, yaitu Metode Elbow (Inertia/WCSS), Skor Silhouette, Indeks Davies-Bouldin, dan Indeks Calinski-Harabasz dalam rentang $K = 2$ sampai $K = 10$. Evaluasi ini dilakukan terhadap algoritma K-Means dan K-Medoids secara bersamaan untuk mendapatkan gambaran perbandingan kinerja.



Gambar 6. Grafik Elbow Method

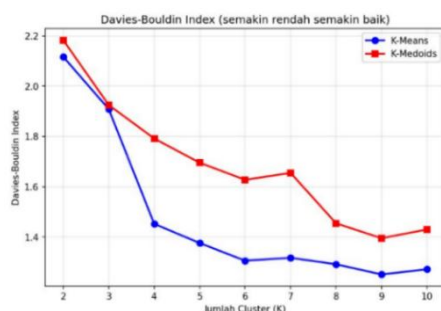
Kurva inertia dari kedua algoritma menunjukkan penurunan monoton tanpa titik "siku" (elbow) yang tajam dan definitif. K-Means secara konsisten menghasilkan nilai inertia yang lebih

rendah dibandingkan K-Medoids pada semua nilai K. Tidak adanya elbow yang jelas mengindikasikan bahwa data tidak memiliki struktur cluster yang sangat tegas secara alami, sehingga metrik tambahan diperlukan untuk mendukung keputusan.



Gambar 7. Grafik Silhouette Score

Evaluasi menggunakan Silhouette Score memperlihatkan perbedaan karakteristik antara kedua metode. Pada K-Means, nilai silhouette cenderung meningkat seiring bertambahnya jumlah klaster dan mencapai nilai tertinggi pada $K = 10$ (0,2078). Sementara itu, K-Medoids menunjukkan fluktuasi nilai dengan puncak pada $K = 9$ (0,1874). Secara umum, K-Means menghasilkan nilai Silhouette Score yang lebih tinggi pada seluruh variasi K.

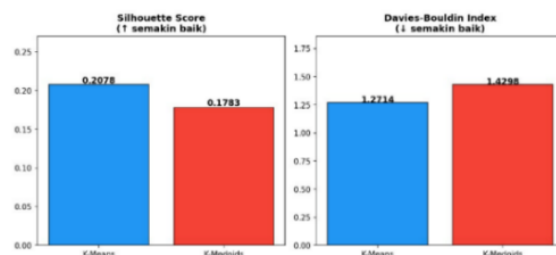


Gambar 8. Grafik Davies-Bouldin Index

Pada metrik Davies-Bouldin Index (nilai lebih rendah menunjukkan kualitas cluster lebih baik), K-Means menunjukkan penurunan nilai yang konsisten hingga mencapai 1,2714 pada $K = 10$, sementara K-Medoids mencapai 1,4298. K-Means unggul di seluruh nilai K yang diuji pada metrik ini. Berdasarkan hasil dari ketiga metrik evaluasi, $K = 10$ ditetapkan sebagai jumlah cluster optimal untuk kedua algoritma, terutama didasarkan pada Silhouette Score K-Means yang mencapai nilai tertinggi pada $K = 10$.

3. Perbandingan Metrik Evaluasi pada $K = 10$

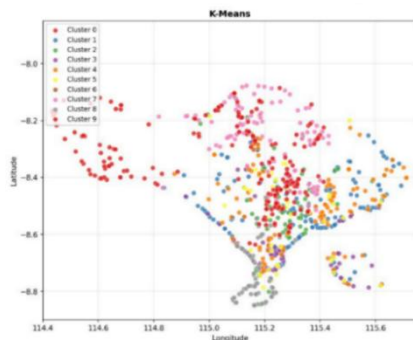
Perbandingan performa kedua algoritma pada $K = 10$ disajikan pada gambar berikut ini.



Gambar 9. Perbandingan Metrik Clustering

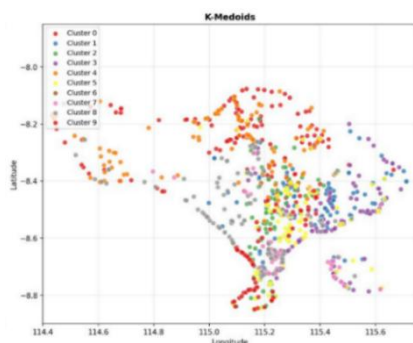
K-Means menunjukkan kinerja yang lebih unggul dibandingkan K-Medoids pada seluruh metrik evaluasi. Peningkatan Silhouette Score sebesar 16,5%, penurunan DBI sebesar 11,1%, dan penurunan MSE sebesar 12,9% menunjukkan bahwa K-Means menghasilkan struktur klaster yang lebih kompak, lebih terseparasi, dan lebih homogen pada data destinasi wisata Bali.

4. Distribusi Geografis Hasil Clustering



Gambar 9. Distribusi Cluster K-Means

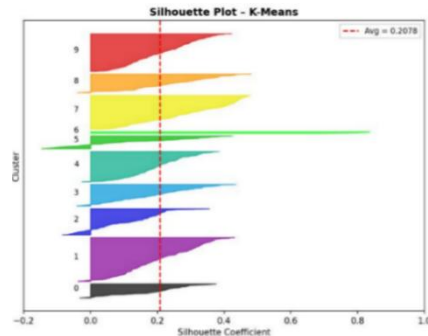
Hasil K-Means menunjukkan distribusi spasial yang cukup konsisten, di mana Cluster 0 didominasi wilayah Jembrana (90%), Cluster 8 dan 9 berada di Badung, serta Cluster 7 terkonsentrasi di Buleleng. Pola ini mengindikasikan bahwa faktor geografis berperan kuat dalam pembentukan cluster oleh K-Means.



Gambar 10. Distribusi Cluster K-Medoids

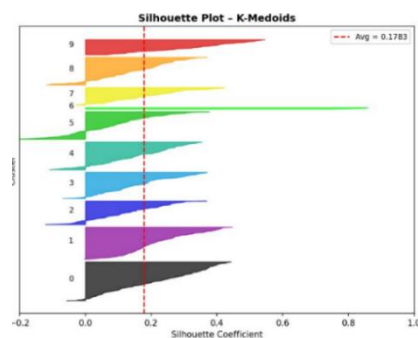
K-Medoids menghasilkan pola sebaran cluster yang lebih tersebar antar wilayah. Hal ini menunjukkan bahwa pengelompokan lebih ditentukan oleh kesamaan atribut non-spasial seperti rating dan kategori, bukan kedekatan geografis.

5. Analisis Silhouette per Sampel



Gambar 11. Silhouette Score K-Means

Pada K-Means (Silhouette rata-rata 0,2078), Cluster 8 dan 9 berada di atas rata-rata dengan pemisahan yang lebih jelas. Cluster 5 memiliki beberapa titik dengan Silhouette sangat tinggi ($>0,8$), sedangkan Cluster 0 menunjukkan cukup banyak nilai negatif yang mengindikasikan adanya ketidaktepatan penempatan data dalam cluster tersebut.



Gambar 12. Silhouette Score K-Medoids

Pada K-Medoids (Silhouette rata-rata 0,1783), Cluster 6 menunjukkan beberapa nilai mendekati 1,0 yang mengarah pada kelompok sangat kecil dan terisolasi. Secara keseluruhan, lebih banyak cluster berada di bawah rata-rata Silhouette dibandingkan K-Means.

6. Distribusi Cluster per Kabupaten/Kota

Distribusi klaster berdasarkan kabupaten/kota disajikan melalui heatmap untuk menunjukkan pola penyebaran secara spasial.



Gambar 13. Heatmap Distribusi Cluster

Distribusi K-Means memperlihatkan pola regional yang mencolok, dengan Cluster 0 didominasi oleh Jembrana (90%), Cluster 7 oleh Buleleng (66 destinasi), dan Cluster 8 oleh Badung (75 destinasi). Ini menunjukkan bahwa K-Means membangun kelompok yang sesuai dengan jarak geografis

4.2. Pembahasan

Hasil pengujian yang dilakukan membuktikan K-Means lebih unggul daripada K-Medoids dalam mengelompokkan destinasi wisata Provinsi Bali berdasarkan seluruh metrik evaluasi yang digunakan. K-Means menghasilkan Silhouette Score sebesar 0,2078, Davies-Bouldin Index sebesar 1,2714, dan MSE sebesar 2,2764, yang keseluruhannya lebih baik dibandingkan K-Medoids. Hasil visualisasi spasial menunjukkan bahwa K-Means membentuk cluster yang lebih terstruktur dibanding K-Medoids. Wilayah

Jembrana dan Buleleng cenderung terkonsentrasi dalam cluster tertentu, sedangkan Badung dan Gianyar tersebar ke beberapa cluster. Hal ini menunjukkan kemampuan K-Means dalam menangkap pola multidimensi data, sehingga menghasilkan cluster yang lebih homogen dan lebih terpisah antar kelompok.

Dari perspektif analisis, perbedaan kinerja K-Means dan K-Medoids dalam penelitian ini berkaitan dengan karakteristik data yang digunakan. Nilai Silhouette Score K-Means sebesar 0,2078 berada pada kategori lemah menurut interpretasi Rousseeuw^[14], yang menunjukkan bahwa meskipun lebih unggul dari K-Medoids, struktur klaster secara umum masih belum kuat. Temuan ini selaras dengan Elbow Method yang tidak menunjukkan titik siku yang tegas, sehingga mengindikasikan bahwa data destinasi wisata cenderung bersifat kontinu tanpa batas klaster yang jelas. Pada Davies-Bouldin Index, K-Means mencatat nilai lebih rendah (1,2714) dibandingkan K-Medoids (1,4298), yang menunjukkan pemisahan klaster yang lebih baik. Sementara itu, MSE yang lebih kecil pada K-Means (2,2764) mengindikasikan centroid yang lebih representatif terhadap anggota klaster dibandingkan medoid. Kondisi ini juga dipengaruhi oleh data yang telah ditransformasi logaritmik, sehingga distribusinya lebih stabil dan tidak terlalu dipengaruhi outlier ekstrem.

Temuan penelitian ini konsisten dengan beberapa studi terdahulu yang membandingkan K-Means dan K-Medoids pada berbagai domain. Siregar et al. melaporkan bahwa K-Means lebih unggul pada data numerik kontinu^[6], yang sejalan dengan karakteristik data destinasi wisata multidimensi pada penelitian ini. Selain itu, Sucahyo et al. menyatakan bahwa K-Means cenderung lebih efektif pada data tanpa outlier ekstrem, sedangkan K-Medoids lebih sesuai untuk data dengan penciran yang signifikan^[7]. Hal ini mendukung peran transformasi logaritmik pada atribut jumlah ulasan yang digunakan dalam penelitian ini, karena mampu menghasilkan distribusi yang lebih stabil dan mengurangi dominasi outlier. Studi Alngatiq et al. juga menunjukkan bahwa pengelompokan destinasi wisata berbasis atribut kuantitatif menghasilkan klaster yang dapat diinterpretasikan secara praktis^[1], sehingga memperkuat relevansi pendekatan yang digunakan. Meskipun demikian, penelitian ini memiliki beberapa keterbatasan. Dataset yang digunakan bersifat statis sehingga tidak mencerminkan dinamika waktu maupun tren kunjungan wisata. Selain itu, variabel yang digunakan masih terbatas pada enam atribut sehingga belum mencakup aspek lain seperti biaya, fasilitas, aksesibilitas, dan sentimen ulasan. Nilai Silhouette Score yang relatif rendah ($<0,3$) pada kedua algoritma menunjukkan

bahwa pemisahan klaster belum kuat, sehingga interpretasi hasil perlu dilakukan dengan hati-hati. Penelitian ini juga belum mempertimbangkan pengaruh inisialisasi acak pada K-Means terhadap stabilitas hasil clustering. Keterbatasan tersebut membuka peluang penelitian lanjutan melalui penambahan data temporal, perluasan fitur, serta eksplorasi metode clustering yang lebih adaptif terhadap kompleksitas data pariwisata

5. Kesimpulan

Pengujian menunjukkan bahwa metode yang digunakan efektif untuk membandingkan kinerja K-Means dan K-Medoids dalam pengelompokan 1.000 destinasi wisata di Provinsi Bali berdasarkan enam atribut multidimensi. Evaluasi menggunakan Silhouette Score, Davies-Bouldin Index, dan MSE dengan jumlah klaster optimal $K=10$ menunjukkan bahwa K-Means konsisten lebih unggul, dengan Silhouette Score 0,2078, DBI 1,2714, dan MSE 2,2764, dibandingkan K-Medoids yang masing-masing sebesar 0,1783, 1,4298, dan 2,6146. Visualisasi spasial juga menunjukkan bahwa hasil clustering K-Means lebih sesuai dengan pola geografis destinasi wisata di Bali sehingga lebih mudah diinterpretasikan secara praktis.

6. Daftar Pustaka

- [1] U. Alngatiq, H. Wamulkan, N. W. Utami, and I. N. Y. Anggara, "Bali Tourist Visits Clustered Via Tripadvisor Reviews Using K-Means Algorithm," vol. 19, no. 2, 2023, doi: 10.33480/pilar.v19i2.4571.
- [2] A. S. Muhammad Qusyairi, Zul Hidayatullah, "Penerapan K-Means Clustering Dalam Pengelompokan Prestasi Siswa Dengan Optimasi Metode Elbow," vol. 7, no. 2, pp. 500–510, 2024, doi:<https://dx.doi.org/10.29408/jit.v7i2.26375>
- [3] A. Fauzan, A. Novianti, R. Rara, and M. Ayu, "Analysis of Hotels Spatial Clustering in Bali : Density- Based Spatial Clustering of Application Noise (DBSCAN) Algorithm Approach," vol. 3, no. 1, pp. 25–38, 2022, doi: 10.20885/EKSAKTA.vol3.iss1.art.
- [4] M. A. J. Amri Muliawan Nur, Hariman Bahtiar, "Implementasi Algoritma K-Means Clustering Dalam Mengelompokkan Kepatuhan Wajib Pajak Bumi dan Bangunan Dengan Optimasi Elbow," vol. 8, no. 1, 2025, doi:<https://dx.doi.org/10.29408/jit.v8i1.27975>
- [5] E. Xiao, "Comprehensive K-Means Clustering," pp. 146–159, 2024, doi: 10.4236/jcc.2024.123009.
- [6] F. M. Siregar, U. Juwardi, and A. K. Hidayah, "Comparative Analysis Of K Means And K Medoids Algorithms In Determining Social Assistance In Padang Sidempuan City , North Sumatra," no. 1, pp. 1–7, 2024, doi:<https://doi.org/10.53697/jkomitek.v4i1.1795>
- [7] F. R. Sucahyo, I. H. Santi, M. F. Rahmat, and D. Fahrizal, "Comparing K-Means and K-medoids algorithms for clustering hamlet regions by tax liabilities in tax determination documents," 2025, doi:<https://doi.org/10.53771/ijstra.2025.8.1.0023>
- [8] A. S. Yaro, H. Bello, E. E. Agbon, A. D. Usman, and A. Abdulaziz, "Comparative Analysis of K-Means and K-Medoids Clustering for Improved Audio-Based Time Difference of Arrival Fingerprinting Systems," vol. 32, no. 1, 2025.
- [9] S. F. Intan, W. Elvira, S. Rahayu, and N. Nurfadilla, "Comparison of the K-Means and K-Medoids Algorithms for Grouping Student Expenditures" pp. 35–40, 2023.
- [10] N. T. Luchia, H. Handayani, and F. S. Hamdi, "Comparison of K-Means and K-Medoids on Poor Data Clustering in Indonesia" vol. 2, no. October, pp. 35–41, 2022.
- [11] A. A. Wani, "Comprehensive analysis of clustering algorithms : exploring limitations and innovative solutions," 2024, doi: 10.7717/peerj-cs.2286.
- [12] N. W. Wardani, P. Gede, S. Cipta, and G. S. Mahendra, "Implementasi Naïve Bayes Pada Data Mining Untuk Mengklasifikasikan Penjualan Barang Terlaris Pada Perusahaan Ritel," vol. 12, no. 3, pp. 656–668, 2023, doi:<https://doi.org/10.23887/jstundiksha.v12i3.38605>
- [13] P. J. R. Leonard Kaufman, "Clustering By Means Of Medoids."
- [14] Y. Hasan, "Pengukuran Silhouette Score dan Davies-Bouldin Index pada Hasil Cluster K-means Dbscan," vol. 06, no. 01, pp. 60–74, 2024.
- [15] R. Indra Herdiana, M Alfin Kamal, Triyani, Mutia Nur Estri, "A more precise elbow method for optimum k-means clustering," pp. 1–22.