Journal of Computing and Informatics Research
ISSN 2808-375X (Media Online)
Vol 4, No 3, July 2025
Hal 354−361
DOI 10.47065/comforch.v4i3.2079

Klasterisasi Lokasi Wisata di Indonesia dengan Menggunakan Algoritma *Hierarchical Clustering*

Puderta Sinulingga[1], Nora Irawani Siregar[2], Andre Hasudungan Lubis[3,*]

[1, 2, 3] Fakultas Teknik, Teknik Informatika, Universitas Medan Area, Medan, Indonesia
Email: [1]pudertalingga2202@gmail.com, [2]norairawani014@gmail.com, [3,*]andrelubis2201@gmail.com
Email Penulis Korespondensi: andrelubis2201@gmail.com

Abstrak−Penelitian ini bertujuan untuk mengelompokkan lokasi wisata di Indonesia berdasarkan rating dan jumlah ulasan menggunakan algoritma *Hierarchical Clustering*. Permasalahan yang diangkat adalah ketimpangan popularitas antar destinasi wisata yang menyebabkan tidak meratanya promosi dan pengelolaan potensi pariwisata di berbagai wilayah. Data sekunder diambil dari platform Kaggle yang mencakup informasi tentang nama tempat, rating, jumlah ulasan, kota, dan provinsi sebagai atribut utama. Tahapan penelitian meliputi pengumpulan data, praproses data (termasuk normalisasi), analisis eksploratif, implementasi algoritma klasterisasi, dan evaluasi hasil secara menyeluruh. Metode *Ward Linkage* digunakan dalam pendekatan agglomeratif untuk membentuk klaster yang optimal dan representatif. Evaluasi klasterisasi dilakukan dengan menggunakan *Silhouette Index* (SI) dan *Davies-Bouldin index* (DBI) sebagai metrik utama yang efektif mengukur kualitas klaster dan konsistensi hasil pengelompokan. Hasil menunjukkan bahwa pemilihan jumlah klaster sebanyak 8 menghasilkan performa terbaik dengan nilai DBI sebesar 0,2606 dan SI sebesar 0,8008. Temuan ini menunjukkan bahwa *Hierarchical Clustering* efektif dalam mengelompokkan destinasi wisata berdasarkan karakteristik kualitas dan popularitas, serta memberikan dasar yang kuat untuk strategi promosi pariwisata yang lebih terarah dan merata di Indonesia, sekaligus mendukung pengembangan pariwisata berkelanjutan dan pemerataan ekonomi.

Kata Kunci: Klasterisasi; *Hierarchical Clustering*; Pariwisata; Indonesia; *Machine Learning*

Abstract−*This study aims to cluster tourist locations in Indonesia based on ratings and number of reviews using the Hierarchical Clustering algorithm. The problem raised is the inequality in popularity between tourist destinations which causes uneven promotion and management of tourism potential in various regions. Secondary data is taken from the Kaggle platform which includes information on place names, ratings, number of reviews, city, and province as the main attributes. The research stages include data collection, data preprocessing (including normalization), exploratory analysis, implementation of the clustering algorithm, and overall evaluation of the results. The Ward Linkage method is used in an agglomerative approach to form optimal and representative clusters. Clustering evaluation is carried out using the Silhouette Index (SI) and Davies-Bouldin index (DBI) as the main metrics that effectively measure cluster quality and consistency of clustering results. The results show that selecting the number of clusters as many as 8 produces the best performance with a DBI value of 0.2606 and an SI of 0.8008. These findings demonstrate that Hierarchical Clustering is effective in grouping tourist destinations based on quality and popularity characteristics, providing a strong foundation for a more targeted and equitable tourism promotion strategy in Indonesia, while supporting sustainable tourism development and economic equality.*

Keywords: *Clustering; Hierarchical Clustering; Tourism; Indonesia; Machine Learning*

1. PENDAHULUAN

Pariwisata merupakan salah satu sektor strategis yang baik di Indonesia terkhusus dalam perekonomian negara, dibuktikan dengan kontribusi signifikan terhadap Produk Domestik Bruto (PBD. Kawasan Asia Tenggara merupakan kawasan yang memiliki keanekaragaman budaya, keindahan alam serta warisan sejarah yang sangat beragam sehingga menjadikan Indonesia sebagai destinasi wisata unggul. Potensi besar tersebut belum sepenuhnya dimanfaatkan secara merata dikarenakan terbatasnya pengelompokan lokasi pariwisata dengan baik. Pengelompokan dapat diatasi dengan klasterisasi menggunakan algoritma tertentu untuk mengelompokkan lokasi wisata berdasarkan karakteristik yang diinginkan [1],[2].

Namun belum ada sistem klasterisasi yang optimal dalam mengatasi permasalahan pengelompokan lokasi-lokasi wisata tersebut berdasarkan katakteristik, fasilitas dan juga lokasi-lokasi yang terfavorit di beberapa daerah. Sehingga menyebabkan ketidak seimbangnya dalam pengenalan lokasi wisata yang ada, dengan itu menyebabkan ada lokasi yang populer sementara beberapa lokasi tidak begitu dikenal oleh banyak orang. Namun disamping itu terdapat tantangan yang harus dihadapi seperti pengelompokan dalam promosi, pengelolaan, dan juga perencanaan untuk kedepannya. Oleh karena itu, diperlukan metode berbasis data seperti *Hierarchical Clustering* yang dapat mengidentifikasi klaster lokasi secara representatif dan terstruktur dengan dua pendekatan yaitu *agglomeratived* dan *divisive* [3] .

Dari beberapa penelitian terdahulu yang salah satunya dari mengatakan Penelitian sistem rekomendasi wisata di Malang Raya menggunakan algoritma *K-Means* menunjukkan beberapa keterbatasan, seperti ketergantungan pada kualitas data, sensitivitas terhadap inisialisasi *centroid*, dan kesulitan dalam menangani outlier serta dinamika waktu. Evaluasi menggunakan *Silhouette Coefficient* menunjukkan bahwa jumlah klaster yang lebih tinggi cenderung memberikan hasil *Clustering* yang lebih baik, terutama di Kota Malang dan Kota Batu. Meskipun algoritma ini efektif dalam mengelompokkan destinasi wisata berdasarkan berbagai faktor, peningkatan pada kualitas data dan metode klasterisasi alternatif masih diperlukan untuk hasil yang lebih optimal [4].

Namun penelitian Agustina pada tahun 2023 mengatakan penelitian ini memiliki keterbatasan dalam hal data yang mungkin tidak mencakup seluruh objek wisata, yang dapat mempengaruhi hasil klasterisasi. Algoritma *Jaccard Similarity Coefficient* hanya mempertimbangkan kesamaan tanpa memperhitungkan bobot atau relevansi komponen. Meskipun demikian, algoritma ini berhasil mengelompokkan objek wisata menjadi 5 klaster berdasarkan nilai kesamaan di atas 0,75, menunjukkan efektivitasnya dalam klasifikasi .

Journal of Computing and Informatics Research
ISSN 2808-375X (Media Online)
Vol 4, No 3, July 2025
Hal 354−361
DOI 10.47065/comforch.v4i3.2079

Untuk penelitian Habiballoh pada tahun 2024 mengatakan penelitian *Clustering* wisata di Jawa Barat menggunakan data terbatas dari satu tahun dan algoritma *K-Means* yang sensitif terhadap pemilihan jumlah klaster, sehingga hasilnya mungkin kurang merefleksikan dinamika wisata jangka panjang dan berisiko bias. Evaluasi *menggunakan Davies Bouldin Index* menunjukkan model dengan tiga klaster cukup efektif, meski keterbatasan dalam validasi eksternal dan asumsi klaster bulat dapat memengaruhi akurasi. Rekomendasi pengembangan pariwisata diberikan berdasarkan klaster, dengan saran untuk penelitian lanjutan yang memperbaiki kualitas data, metode *Clustering*, dan mempertimbangkan perubahan temporal.

Lalu Dalam penelitian yang berjudul Mapping Domestic and Foreign Tourists in East Java Using C-Means Clustering oleh Marita Qori'atunnadyah (2024) ditemukan bahwa kabupaten/kota di Jawa Timur terbagi menjadi tiga klaster kunjungan wisatawan domestik dan lima klaster untuk wisatawan asing berdasarkan metode klasterisasi C-Means [5].

Dan penelitian Seftia pada tahun 2024 menggunakan algoritma *K-Means* untuk mengelompokkan data kunjungan wisata, namun terbatas oleh ukuran sampel kecil, ketergantungan pada pemilihan *centroid* awal, dan tidak mempertimbangkan variasi musiman. Algoritma ini berhasil membagi data menjadi dua klaster yang berbeda berdasarkan jumlah pengunjung, dengan performa yang cukup baik namun memiliki keterbatasan pada pengelompokan data yang kompleks. Untuk hasil yang lebih akurat dan aplikatif, diperlukan evaluasi metrik tambahan serta strategi optimasi dan pengembangan lebih lanjut dalam penelitian selanjutnya [6].

Oleh karena itu, untuk mengatasi tantangan dalam pengelompokan lokasi wisata yang ada di Indonesia dengan karakteristik yang beragam, algoritma *Hierarchical Clustering* dapat menjadi solusi yang efektif dibanding algoritma-algoritma terdahulu. Hal ini dikarenakan algoritma Hierarchical clufstering mampu mengelompokkan destinasi wisata berdasarkan fitur spasial dan lebih representatif lagi, sehingga sangat mendukung dalam promosi pariwisata yang lebih efektif. Pengelola sumber daya wisata juga dapat dioptimalkan sehingga dapat meningkatkan pemerataan pengunjung dan efektif dalam pengembangan pariwisatadi Indonesia. Oleh karena itu, implementasi *Hierarchical Clustering* menjadi solusi utama yang diusulkan untuk meningkatkan kualitas klasterisasi lokasi wisata di Indonesia, serta memberikan dasar yang kuat untuk pengambilan keputusan strategis dalam pengembangan sektor pariwisata nasional.

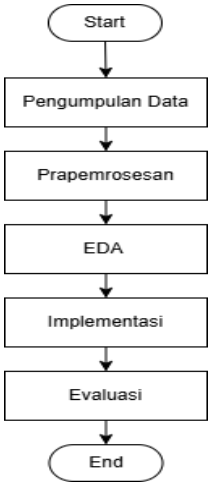
2. METODE PENELITIAN

2.1 Wisata

Wisata adalah kegiatan perjalanan yang dilakukan oleh seseorang atau sekelompok orang ke suatu tempat di luar lingkungan tempat tinggalnya untuk tujuan rekreasi, pendidikan, spiritual, kesehatan, atau kepentingan lainnya dalam jangka waktu sementara wisara. Kegiatan wisata biasanya melibatkan unsur perjalanan, tujuan yang menarik, serta pengalaman baru yang memberikan kepuasan bagi wisatawan. Dalam konteks yang lebih luas, wisata juga mencakup seluruh aktivitas yang mendukung perjalanan tersebut, seperti transportasi, akomodasi, kuliner, dan jasa pemandu wisata [7],[8],[9].

2.2 Tahapan Penelitian

Setiap langkah tentunya memiliki peran dan fungsi yang saling berkaitan untuk mencapai tujuan yang jelas. Oleh karena itu, proses yang sistematis sangat penting sangat penting untuk menghasilkan data yang valid dan real. Sehingga tahapan penelitian ini disusun secara runtut dan terstruktur mulai dari pengumpulan data sampai evaluasi model. Untuk melihat lebih jelas, bisa dilihat dari Gambar 1 diagram alir dibawah.



Gambar 1. Diagram Alir

Tahapan pertama yaitu tahapan Pengumpulan data, dimana data ini diperoleh dari mengakses situs web Kaggle yang menyediakan berbagai dataset. Data yang dipakai merupakan data sekunder yang sudah tersedia dan relevan dengan

Journal of Computing and Informatics Research
ISSN 2808-375X (Media Online)
Vol 4, No 3, July 2025
Hal 354−361
DOI 10.47065/comforch.v4i3.2079

permasalahan yang sedang diteliti. Data ini diakses dari situs https://www.kaggle.com/datasets/masdarulrizqi/tourism-data-on-java. Adapun rangkuman datanya dapat dilihat pada Tabel 1 di bawah ini.

Tabel 1. Data Asli

Nama Tempat	Rating	Total *Review*	Kota	Provinsi
Alas Purwo National Park	4.50	1.449	Banyuwangi	Jawa Timur
Green Bay	4.70	2.322	Banyuwangi	Jawa Timur
De Djawatan Forest	4.60	6.906	Banyuwangi	Jawa Timur
Banyuwangi Park	4.50	1.242	Banyuwangi	Jawa Timur
Tourism Village Osing	4.20	1.604	Banyuwangi	Jawa Timur
...	...	...	...	...
Telaga Biru Cisoka	4	6.286	Tangerang	Banten
Danau Barat Alam Sutera	4.50	153	Tangerang	Banten
Koja Cliff Park	4	1.690	Tangerang	Banten
Flying Deck Cisadane	4.50	438	Tangerang	Banten
Situ Batusari	4.60	340	Tangerang	Banten

Selanjutnya yaitu tahapan Prapemrosesan data, tahapan ini berguna untuk memastikan data dalam kondisi yang siap digunakan dalam penelitian (Md et al., 2023). Pada tahapan ini, biasanya akan dilakukan pemeriksaan terhadap *missing value* atau bisa dikatakan nilai yang hilang pada dataset serta penanganannya agar tidak perlu dianalisis guna untuk mengemat waktu. Selain itu juga, ada *feature engineering* untuk memodifikasi fitur yang ada agar lebih representatif terhadap masalah yang dihadapi. Kemudian data di normalisasi dengan menggunakan metode *Min-Max Scaler* yang mengubah nilai fitur kedalam rentang 0 sampai 1, sesuai dengan rumus berikut :

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

Tahapan berikutnya ialah ***Exploratory Data Analysis*** **(EDA),** dimana pada tahap ini dilakukan analisis statistik deskriptif agar memahami karakteristik data secara umum, lalu pemeriksaan korelasi antar fitur untuk mengidentifikasi hubungan yang mungkin ada antar variabel, serta visualisasi data dengan berbagai plot untuk memudahkan interpretasi pola dan distribusi dalam melihat dan mengalisis data.

Selanjutnya ialah tahapan Implementasi, yang dilakukan dengan menerapkan algoritma *machine learning* pada kasus yang diteliti. Dalam penelitian ini, peneliti memilih algoritma *Hierarchical Clustering*.

Metode *hirarchical Clustering* adalah metode klaster yang dilakukan secara bertahap (hirarki) dengan menggunakan konsep penggabungan dua klaster kecil yang memiliki jarak terdekat menjadi satu klaster yang lebih besar yang disebut dengan metode *Agglomeratif Hierarchical Clustering* atau pemecahan klaster besar ke beberapa klaster yang lebih kecil dengan dasar ketidakmiripan paling tinggi klaster-klaster lebih kecil yang disebut dengan metode *Divisive*. Pembentukan anggota klaster pada metode hirarki menggunakan bagan atau dendogram.

Terdapat dua prosedur pada metode *Hirearchical*, yaitu prosedur aglomeratif dan prosedur divisive. Dalam penelitian ini, akan digunakan salah pembagian metode algomeratif, yaitu Metode *Ward* dengan jarak antar dua *cluster* adalah total jumlah kuadrat dua *cluster* pada masing masing *variable*. Metode ini berbeda dengan metode lainnya karena menggunakan pendekatan analisis varians untuk menghitung jarak antar *cluster* atau metode ini meminimumkan jumlah kuadrat (ESS) (Pada et al., 2024). Algoritma metode hirarki agglomeratif secara umum untuk mengelompokkan N objek dapat dilihat dari rumus berikut.

$$ESS = \sum_{j=1}^{n} x_j^2 - \frac{1}{n}\left(\sum_{j=1}^{n} x_j\right)^2 \quad (2)$$

Pseudocode Hierarcical Clustering
Input: Data points D = $\{x_1, x_2, x_3, \ldots x_n\}$
Output: Dendrogram / Hierarchical cluster tree

1. Mulai dengan setiap titik data sebagai *cluster* sendiri (*n cluster*)
2. Hitung matriks jarak antar semua *cluster* (berdasarkan jarak antar titik)
3. WHILE jumlah *cluster* > 1 DO
   a. Temukan dua *cluster* terdekat A dan B berdasarkan metode *linkage* yang dipilih
   b. Gabungkan *cluster* A dan B menjadi *cluster* baru C
   c. Update matriks jarak
   d. Hitung jarak antara C dan setiap *cluster* lain menggunakan rumus *linkage*
   e. Hapus baris dan kolom matriks jarak untuk *cluster* A dan B
   f. Tambahkan baris dan kolom untuk *cluster* C
4. Kembalikan *dendrogram* atau hierarki *cluster*.

Lalu pada tahapan terakhir terdapat tahaham **Evaluasi,** yang artinya data yang telah diimplementasi akan diuji atau dievaluasi dengan menggunakan validasi SI *(Silhoutte Index)* dan DBI *(Davis Bouldin Index).*

SI *(Silhoutte Index)*

Journal of Computing and Informatics Research
ISSN 2808-375X (Media Online)
Vol 4, No 3, July 2025
Hal 354−361
DOI 10.47065/comforch.v4i3.2079

Silhouette Index menghitung rata-rata masing-masing titik pada sekumpulan data. Adapun rumus yang digunakan untuk menghitung *Silhouette Index* adalah [10],[11],[12].

$$SI = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{b(i)-a(i)}{\max\{a(i),b(i)\}}\right) \quad (3)$$

DBI *(Davis Bouldin Index)* merupakan sebuah metode evaluasi internal *cluster*, yang dimana hasil baik tidaknya hasil *cluster* dapat dilihat dari kohesi dan sparasi. Kohesi adalah besar jarak antar data yang bertujuan untuk mengukur seberapa dekat data-data pada *cluster* yang sama. Sedangkan separasi merupakan pengukuran perbedaan data-data yang ada pada *cluster* yang berbeda. Rumus DBI dapat dilihat seperti berikut [13],[14],[15]

$$DBI = \frac{1}{k}\sum_{i=1}^{k} max_{i\neq j}\left(R_{i,j,\ldots k}\right) \quad (4)$$

Dimana keterangan dari rumus-rumus diatas, dapat disimpulkan bahwa jika nilai akurasi DBI yang diperoleh mendekati 0 maka akurasi DBI semakin baik. Sebaliknya jika nilai akurasi SI yang diperoleh mendekati 1 maka akurasi SI akan semakin baik [16],[17],[18].

3. HASIL DAN PEMBAHASAN

3.1 Prapemrosesan Data

Dalam tahap prapemrosesan data, variabel kategorikal seperti *City* (kota) dan *Province* (provinsi) perlu diubah menjadi bentuk numerikal agar dapat digunakan dengan efektif dalam algoritma klasterisasi yang berbasis perhitungan jarak. Proses transformasi ini dilakukan dengan menggunakan teknik label encoding, yang secara sistematis memberikan setiap kategori unik pada variabel tersebut representasi angka yang berbeda. Setiap kategori yang ada, seperti nama-nama kota dan provinsi, akan digantikan dengan angka yang secara konsisten mewakili kategori tersebut dalam dataset. Dengan cara ini, data yang pada awalnya bersifat kualitatif dan tidak dapat langsung diproses oleh algoritma berbasis jarak, seperti *Hierarchical Clustering*, dapat diubah menjadi format numerik yang dapat diintegrasikan secara efisien. Transformasi ini memungkinkan algoritma untuk menghitung jarak antar data dan menghasilkan klaster-klaster yang merepresentasikan pola sebaran geografis dari lokasi-lokasi wisata secara lebih akurat. Hasil dari transformasi data ini, yang menunjukkan bagaimana variabel kategorikal tersebut diterjemahkan ke dalam angka, dapat dilihat pada Tabel 2.

Tabel 2. Hasil *Transform* Data

Rating	Total *Review*	*City*	*Province*
4,5	1449	2	3
4,7	2322	2	3
4,6	6906	2	3
4,5	1242	2	3
4,2	37	2	3

Kemudian, Tabel 3 menunjukkan hasil dari normalisasi data yang telah dilakukan pada dataset. Normalisasi data merupakan langkah penting yang harus dilakukan sebelum menjalankan klasterisasi karena setiap variabel harus memiliki kontribusi yang setara dalam perhitungan jarak antar data. Tanpa normalisasi, variabel dengan skala yang lebih besar, seperti harga atau jumlah pengunjung, akan mendominasi perhitungan jarak, sementara variabel dengan skala yang lebih kecil, seperti kategori kota dan provinsi, akan memiliki pengaruh yang lebih kecil. Proses normalisasi ini memastikan bahwa setiap variabel memberikan kontribusi yang proporsional terhadap jarak yang dihitung. Dalam Tabel 3, dapat dilihat bahwa jarak antar nilai pada data yang telah dinormalisasi tidak terlalu jauh, yang menandakan bahwa skala data telah disesuaikan sehingga tidak ada variabel yang mendominasi. Selain itu, nilai-nilai yang memiliki entri yang sama, seperti data dari kota dan provinsi yang sama, memiliki nilai yang serupa, yang juga mencerminkan konsistensi dalam representasi data. Tabel ini merupakan rangkuman yang menggambarkan hasil dari normalisasi data lokasi wisata di Indonesia, yang siap untuk digunakan dalam tahap klasterisasi selanjutnya.

Tabel 3. Hasil Normalisasi Data

Rating	Total *Review*	*City*	*Province*
0.80	0.014844	0.057143	0.75
0.88	0.023793	0.057143	0.75
0.84	0.070786	0.057143	0.75
0.80	0.012722	0.057143	0.75
0.68	0.000369	0.057143	0.75

3.2 EDA (*Exploratory Data Analysis)*

Pada bagian ini, disajikan hasil eksplorasi data awal yang meliputi rangkuman statistik deskriptif dari variabel utama dalam dataset lokasi wisata di Indonesia. Proses eksplorasi data ini bertujuan untuk memberikan gambaran umum yang

Journal of Computing and Informatics Research
ISSN 2808-375X (Media Online)
Vol 4, No 3, July 2025
Hal 354−361
DOI 10.47065/comforch.v4i3.2079

lebih jelas mengenai karakteristik dasar data sebelum dilakukan tahap analisis lebih lanjut. Tabel 4 berikut menyajikan informasi terkait distribusi nilai rating dan jumlah ulasan, yang merupakan dua variabel penting yang dapat memberikan wawasan mendalam tentang kualitas dan popularitas setiap lokasi wisata. Pemahaman mengenai distribusi kedua variabel ini menjadi sangat krusial, karena keduanya akan menjadi dasar dalam menentukan bagaimana klasterisasi dilakukan. Dengan demikian, tabel ini memberikan informasi yang sangat berguna dalam memahami pola sebaran data, yang nantinya akan mempengaruhi hasil klasterisasi dan interpretasi hasil analisis.

Tabel 4. Statistik Data

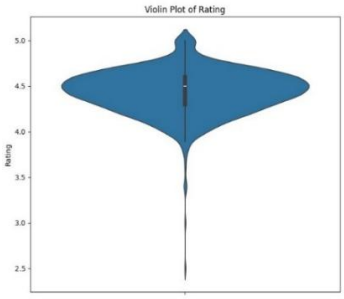
Statistik	Rating	Total *Review*
count	1017.000000	1017.000000
mean	4.437070	3737.595870
std	0.243159	9258.864885
min	2.500000	1.000000
25%	4.300000	174.000000
50%	4.500000	713.000000
75%	4.600000	3023.000000
max	5.000000	97549.000000

Tabel 4 menjelaskan rangkuman statistik dari 1.017 lokasi wisata yang berada di Indonesia, berdasarkan dua variabel utama, yaitu rating dan total *review*. Variabel Rating menunjukkan hasil yang cenderung standar, dengan penilaian terhadap sebagian besar lokasi wisata yang cenderung tinggi dan tidak terlalu bervariasi. Hal ini menunjukkan bahwa mayoritas tempat wisata di Indonesia mendapatkan penilaian yang baik dari pengunjung, meskipun terdapat sedikit variasi di antara nilai-nilai rating tersebut. Hasil ini juga saling berkaitan dengan nilai plot yang ditampilkan di bawah, yang memberikan gambaran visual mengenai konsistensi rating di seluruh lokasi wisata. Di sisi lain, variabel total *review* menunjukkan hasil yang jauh lebih tersebar dan tidak merata, dengan sebagian besar lokasi wisata hanya mendapatkan sedikit ulasan, sementara beberapa lokasi lainnya mendapatkan perhatian yang sangat besar dari pengunjung. Perbedaan yang mencolok ini mengindikasikan bahwa hanya sejumlah kecil lokasi wisata yang berhasil menarik perhatian banyak pengunjung dan mendapatkan jumlah ulasan yang signifikan. Hal ini menunjukkan adanya ketidakseimbangan dalam popularitas, di mana sebagian besar lokasi wisata mungkin kurang dikenal atau hanya dikunjungi oleh segelintir orang, sementara lokasi wisata tertentu mendapatkan lebih banyak perhatian dari publik. Berikut ini pada Tabel 5 merupakan Kolerasi Data

Tabel 5. Kolerasi Data

Korelasi	Rating	Total *Review*	*City*	*Province*
rating	1.00	0.15	-0.02	0.07
total_*Review*s	0.15	1.00	-0.00	0.03
City	-0.02	-0.00	1.00	-0.12
Province	0.07	0.03	-0.12	1.00
rating	1.00	0.15	-0.02	0.07

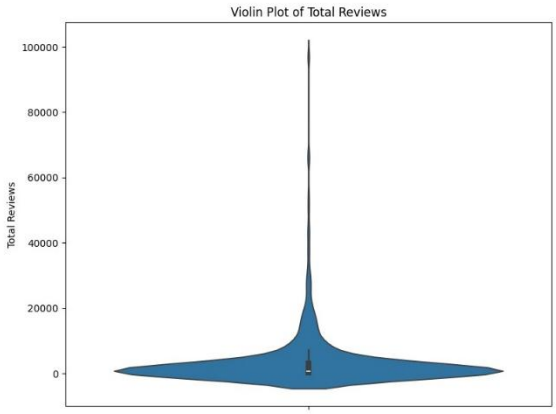
Tabel 5 tentang korelasi di atas menunjukkan bahwa lokasi-lokasi wisata di Indonesia tidak memiliki hubungan korelasi yang baik, yang berarti tidak terdapat hubungan linier yang kuat antara variabel-variabel seperti rating, total *review*, kota, maupun provinsi. Namun, dari tabel di atas, terlihat bahwa rating dengan total *review* memiliki korelasi yang cukup signifikan yaitu sekitar 0,15, yang menunjukkan adanya kecenderungan bahwa semakin tinggi rating suatu tempat, maka jumlah ulasannya juga cenderung lebih banyak. Pada korelasi total *review* dengan rating juga menunjukkan nilai yang sama, yaitu 0,15, memperkuat temuan sebelumnya. Untuk korelasi berikutnya, yaitu antara variabel *city* dengan variabel lain, tidak menunjukkan korelasi yang baik karena nilai korelasinya negatif, yang mengindikasikan hubungan yang berlawanan arah atau tidak saling terkait secara linier. Selanjutnya, korelasi antara provinsi dengan rating menunjukkan nilai korelasi yang cukup baik, menandakan bahwa lokasi provinsi mungkin memiliki pengaruh terhadap rating yang diberikan penggguna, berikut dapat dilihat pada Gambar 2.



Gambar 2. Plot Rating

Journal of Computing and Informatics Research
ISSN 2808-375X (Media Online)
Vol 4, No 3, July 2025
Hal 354−361
DOI 10.47065/comforch.v4i3.2079

Gambar 2 menunjukkan gambar violin plot yang menggambarkan distribusi rating dari data lokasi wisata di Indonesia. Berdasarkan gambar tersebut, dapat dilihat bahwa sebagian besar lokasi wisata di Indonesia memiliki rating yang sangat baik, yang berkisar pada angka 4,5. Hal ini menunjukkan bahwa mayoritas destinasi wisata di Indonesia berhasil memberikan pengalaman yang memuaskan bagi pengunjungnya, sehingga mendapatkan penilaian positif yang tinggi. Rating yang konsisten tinggi ini mencerminkan kualitas pelayanan, fasilitas, dan pengalaman yang diberikan oleh destinasi wisata tersebut. Dengan demikian, sebagian besar tempat wisata di Indonesia telah berhasil memenuhi ekspektasi pengunjung dan memberikan nilai lebih yang memperkuat citra positif destinasi wisata di negara ini.
Namun, plot ini juga menggambarkan adanya beberapa lokasi wisata yang memiliki rating yang relatif rendah, yang mungkin disebabkan oleh faktor-faktor tertentu seperti fasilitas yang kurang memadai, pelayanan yang tidak memuaskan, atau kurangnya promosi dan eksposur. Rating rendah ini bisa menjadi indikasi adanya masalah dalam pengalaman pengunjung yang perlu diperbaiki, seperti kebersihan, kualitas tempat, atau kurangnya informasi tentang destinasi tersebut. Selain itu, faktor eksternal seperti kondisi cuaca atau ketidaksesuaian antara harapan pengunjung dan kenyataan yang dihadapi saat berkunjung juga dapat mempengaruhi rating yang diberikan. Meskipun demikian, dari keseluruhan gambar, dapat disimpulkan bahwa mayoritas tempat wisata di Indonesia telah berhasil memperoleh nilai yang baik dari para pengunjung. Hal ini mengindikasikan bahwa destinasi wisata di Indonesia, secara umum, dianggap menarik dan memiliki kualitas yang cukup baik, yang dapat meningkatkan daya tarik negara ini sebagai tujuan wisata internasional. Dengan sebagian besar destinasi yang mendapat penilaian positif, Indonesia dapat lebih percaya diri dalam mempromosikan dirinya sebagai destinasi pariwisata unggulan, dengan banyak lokasi yang menawarkan pengalaman yang memuaskan bagi wisatawan. Oleh karena itu, hasil ini menunjukkan potensi besar bagi sektor pariwisata Indonesia untuk terus berkembang, sembari mengatasi area-area yang masih membutuhkan perbaikan untuk memastikan kesuksesan jangka panjang. Selanjutnya pada Gambar 3 dibawah ini merupakan Plot Total Review.



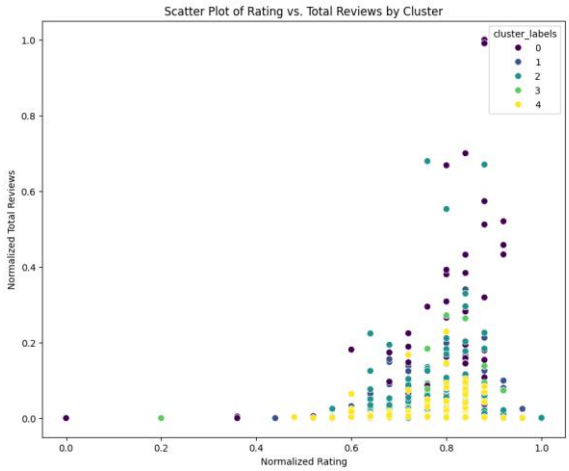
Gambar 3. Plot Total *Review*

Gambar 3 di atas menjelaskan violin plot untuk total *review*, yang menggambarkan distribusi jumlah ulasan yang diterima oleh lokasi wisata di Indonesia. Dari plot ini, dapat dilihat bahwa sebagian besar lokasi wisata memiliki jumlah ulasan yang relatif rendah, yang tercermin dari plot yang melebar di dekat angka 0. Kondisi ini menunjukkan bahwa mayoritas lokasi wisata di Indonesia belum mendapatkan perhatian yang cukup dari pengunjung, atau mungkin kurang terekspos di kalangan wisatawan. Ini tentunya bukan hasil yang optimal karena menunjukkan kurangnya interaksi atau feedback dari pengunjung. Namun, terdapat juga sejumlah lokasi wisata yang memiliki jumlah ulasan yang sangat tinggi, bahkan ada yang mencapai lebih dari 100.000 ulasan. Hal ini menunjukkan adanya beberapa destinasi wisata yang sangat populer dan banyak dikunjungi, sehingga mendapat banyak ulasan dari pengunjung. Ketidakseimbangan ini mengindikasikan bahwa popularitas tempat wisata di Indonesia tidak merata, dengan sebagian kecil lokasi wisata yang mendominasi.

3.3 Implementasi

Gambar 4 dibawah ini menjelaskan tentang visualisasi hasil klasterisasi menggunakan *Hierarchical Clustering* yang ditampilkan dalam bentuk scatter plot, dengan rating pada sumbu X dan total *review* pada sumbu Y. Plot ini menunjukkan bagaimana data terbagi menjadi lima klaster yang berbeda, yang masing-masing mewakili pola yang unik dalam distribusi rating dan total *review*. Sebagian besar titik data terkonsentrasi pada area dengan rating tinggi, namun dengan total *review* yang rendah. Hal ini mengindikasikan bahwa banyak lokasi wisata di Indonesia yang memiliki kualitas yang baik atau sangat baik, namun belum cukup populer di kalangan pengunjung, sehingga jumlah ulasan yang diterima masih terbatas. Selanjutnya, terdapat klaster lain yang menunjukkan lokasi wisata dengan rating tinggi dan total *review* yang juga bagus. Klaster ini menggambarkan lokasi wisata yang tidak hanya memiliki kualitas yang baik tetapi juga telah mendapatkan perhatian yang signifikan dari pengunjung, sehingga dapat dianggap sebagai destinasi wisata yang tidak hanya bagus tetapi juga populer dan banyak dikenal oleh masyarakat. Dengan demikian, scatter plot ini memberikan wawasan yang jelas mengenai distribusi kualitas dan popularitas tempat wisata di Indonesia, serta mengidentifikasi pola-pola tertentu yang bisa dijadikan acuan untuk strategi pengembangan dan promosi destinasi wisata di masa depan.

Journal of Computing and Informatics Research
ISSN 2808-375X (Media Online)
Vol 4, No 3, July 2025
Hal 354−361
DOI 10.47065/comforch.v4i3.2079



Gambar 4. *Implementasi Hierarchical Clustering*

3.4 Evaluasi

Tabel 6 di bawah ini menjelaskan hasil dari penelitian yang telah dilakukan menggunakan Algoritma Hierarchical Clustering untuk mengelompokkan lokasi wisata di Indonesia. Berdasarkan tabel tersebut, dapat dilihat bahwa jumlah cluster yang optimal adalah 12, yang menunjukkan hasil terbaik di antara jumlah cluster lainnya. Jumlah cluster yang optimal ini dipilih berdasarkan evaluasi menyeluruh terhadap berbagai metrik yang mengukur efektivitas dan keakuratan klasterisasi. Hasil evaluasi untuk jumlah cluster 12 ini menunjukkan nilai DBI (Davies-Bouldin Index) sebesar 0.2882 dan SI (Silhouette Index) sebesar 0.8328. Nilai DBI yang lebih rendah mengindikasikan bahwa klaster yang terbentuk lebih terpisah dengan jelas, yang berarti bahwa klaster-klaster tersebut lebih homogen dan tidak tumpang tindih satu sama lain. Semakin rendah nilai DBI, semakin baik pembagian antara klaster, karena menandakan bahwa jarak antar klaster lebih jauh dan klaster-klaster yang ada memiliki kesamaan internal yang tinggi. Sebaliknya, nilai SI yang lebih tinggi, yaitu mendekati 1, menunjukkan bahwa jarak antar titik dalam setiap klaster relatif lebih dekat, dan titik-titik antar klaster cukup terpisah, yang menandakan kualitas klaster yang baik. Nilai SI yang mendekati 1 menandakan bahwa model klasterisasi ini tidak hanya mengelompokkan data dengan efisien, tetapi juga memastikan bahwa data dalam masing-masing klaster cukup homogen dan terpisah dengan jelas dari klaster lainnya. Dengan demikian, akurasi penelitian yang menggunakan jumlah cluster sebanyak 12 sudah tergolong baik, karena nilai DBI yang rendah dan nilai SI yang tinggi menunjukkan bahwa model klasterisasi ini dapat membagi data lokasi wisata di Indonesia secara efektif dan efisien. Klaster-klaster yang dihasilkan tidak hanya memiliki kualitas yang tinggi, tetapi juga memberikan informasi yang lebih jelas dan lebih terstruktur tentang pola-pola yang ada dalam data lokasi wisata, sehingga hasil klasterisasi ini dapat digunakan untuk analisis lebih lanjut dan pengambilan keputusan yang lebih tepat.

Tabel 6. Hasil Evaluasi

N_*cluster*	DBI	SI
6	0.5006	0.7220
7	0.4887	0.6790
10	0.4283	0.7551
11	0.3590	0.7867
12	0.3471	0.8328
8	0.2606	0.8008
15	0.2882	0.8328

4. KESIMPULAN

Penelitian ini membuktikan bahwa algoritma *Hierarchical Clustering* mampu secara efektif mengelompokkan lokasi wisata di Indonesia berdasarkan rating dan jumlah ulasan. Dengan melalui tahapan yang sistematis mulai dari pengumpulan data, normalisasi menggunakan *Min-Max Scaler*, hingga penerapan metode *Ward Linkage*, penelitian ini berhasil membentuk klaster yang representatif terhadap kondisi wisata di Indonesia. Hasil analisis menunjukkan bahwa mayoritas lokasi wisata memiliki rating tinggi namun jumlah ulasan yang rendah, mengindikasikan bahwa banyak destinasi berkualitas masih kurang terekspos. Evaluasi dengan *Silhouette Index* dan *Davies-Bouldin index* menunjukkan bahwa jumlah klaster optimal adalah 8, dengan SI sebesar 0,8008 dan DBI sebesar 0,2606, menandakan pemisahan klaster yang baik. Kelebihan metode ini adalah kemampuannya mengungkap struktur hierarki dari klaster secara alami, sementara keterbatasannya terletak pada kompleksitas perhitungan saat jumlah data besar. Untuk pengembangan lebih lanjut, penelitian dapat ditingkatkan dengan integrasi data spasial, data waktu nyata, serta memperluas cakupan wilayah dan atribut wisata untuk mendukung promosi dan pengelolaan destinasi yang lebih efektif dan merata di seluruh Indonesia. Selain itu, penggabungan metode klasterisasi lain atau pendekatan *Hybrid* dengan deep learning dapat

ISSN 2808-375X (Media Online)
Vol 4, No 3, July 2025
Hal 354−361
DOI 10.47065/comforch.v4i3.2079

dieksplorasi untuk meningkatkan akurasi dan fleksibilitas dalam menangani data yang lebih kompleks dan dinamis. Penambahan fitur seperti faktor musiman, tren pengunjung, dan ulasan teks juga berpotensi memberikan insight lebih dalam untuk segmentasi destinasi wisata yang lebih komprehensif. Dengan demikian, hasil penelitian ini dapat menjadi dasar yang kuat untuk pengambilan keputusan strategis dalam pengembangan pariwisata yang berkelanjutan dan adaptif terhadap perubahan pasar.

REFERENCES

[1] S. C. Sitompul, R. Nadza, F. Sihotang, T. Syahputra, and A. Budiman, "Sistem Pendukung Keputusan Pemilihan Hotel Terbaik Di Kota Medan Menggunakan Metode Edas," *J. Ilmu Komput. Dan Sist. Inf.*, Vol. 4, No. 1, Pp. 68–77, 2025, Doi: 10.70340/Jirsi.V4i1.174.

[2] M. Tinambunan And S. Sintaro, "Aplikasi Restfull Pada Sistem Informasi Geografis Pariwisata Kota Bandar Lampung," *J. Inform. Dan Rekayasa Perangkat Lunak*, Vol. 2, No. 3, Pp. 312–323, 2021.

[3] N. Lestari, A. Ambarita, And R. G. Hafel, "Aplikasi E-Tourism Berbasis Android Sebagai Panduan Dan Media Promosi Objek Pariwisata Di Maluku Utara," *J. Ilm. Ilk. Komput. Inform.*, Vol. 6, No. 1, Pp. 11–20, 2023.

[4] H. Di Kesuma, S. Hamidani, P. Studi, S. Informasi, And U. Palembang, "Optimalisasi Strategi Wisata Di Kota Pagar Alam Menggunakan Algoritma K-Means Clustering Optimization Of Tourism Strategies In Pagar Alam City Using The K- Means Clustering Algorithm," *J. Ilm. Bin. Stmik Bina Nusant. Jaya*, Vol. 5, No. 1, Pp. 86–92, 2023, Doi: 10.52303/Jb.V5i1.102.

[5] M. Qori, "Mapping Domestic And Foreign Tourists In East Java Using C-Means Clustering," *Stat. Dan Apl.*, Vol. 8, No. 1, Pp. 54–62, 2024, Doi: 10.21009/Jsa.

[6] M. Seftia, E. Efan, And A. Arif, "Optimization Of Tourism Strategies In Pagar Alam City Using The K- Means Clustering Algorithm," *J. Inform.*, Vol. 01, No. 02, Pp. 6–11, 2024.

[7] S. B. W. Rizky, "Perbandingan Algoritma K-Means Dan Hierarchical Clustering Dalam Segmentasi Kabupaten/Kota Di Jawa Timur Berdasarkan Data Perjalanan Dan Pergerakan Wisatawan," *Jati (Jurnal Mhs. Tek. Inform.*, Vol. 9, No. 5, Pp. 8499–8506, 2025.

[8] M. Herviany, S. Putri Delima, T. Nurhidayah, And Kasini, "Comparison Of K-Means And K-Medoids Algorithms For Grouping Landslide Prone Areas In West Java Province," *Malcom Indones. J. Mach. Learn. Comput. Sci.*, Vol. 1, No. 1, Pp. 34–40, 2021.

[9] A. Salam, D. Adiatma, And J. Zeniarja, "Implementasi Algoritma K-Means Dalam Pengklasteran Untuk Rekomendasi Penerima Beasiswa Ppa Di Udinus," *Joins (Journal Inf. Syst.*, Vol. 5, No. 1, Pp. 62–68, 2020, Doi: 10.33633/Joins.V5i1.3350.

[10] W. M. Putri, E. Asril, And U. L. Kuning, "Analisis Clustering Buku Sebagai Upaya Untuk Meningkatkan Minat Baca Siswa Pada Perpustakaan Sma Negeri 3 Pekanbaru," *Prosiding-Seminar Nas. Teknol. Inf. Ilmu Komput.*, Vol. 2, No. 1, Pp. 313–323, 2023.

[11] M. V. Alhafiz And P. F. Nuryananda, "Analisis Kesiapan Pengelola Wisata Romokalisari Adventure Land Dalam Menggunakan Platform Digital Wisata," *Jiip-Jurnal Ilm. Ilmu Pendidik.*, Vol. 7, No. 12, Pp. 13615–13623, 2024.

[12] S. Amellia And Y. Kurnia, "Perbandingan Algoritma K-Means Dan Hierarchical Clustering Dalam Implementasi Sistem Rekomendasi Destinasi Wisata Di Indonesia," *Poters (Proceedings Technol. Eng. Comput.*, Vol. 1, No. 2, Pp. 308–315, 2025.

[13] Mohammad Ferdiansyah And Umi Chotijah, "Implementasi Algoritme K-Means++ Untuk Clustering Penjualan Bahan Bangunan," *J. Ilm. Tek. Inform. Dan Komun.*, Vol. 4, No. 1, Pp. 181–193, 2024, Doi: 10.55606/Juitik.V4i1.767.

[14] P. N. P. Artana And E. P. Mandyartha, "Penerapan Data Mining Pada Algotirma Hierarchical Clustering Tentang Pengelolaan Mitra Perjalanan Wisatawan Bali Backpaker," *Jati (Jurnal Mhs. Tek. Inform.*, Vol. 7, No. 4, Pp. 2903–2909, 2023.

[15] N. T. Luchia *Et Al.*, "Analisis Sentimen Ulasan Wisatawan Terhadap Destinasi Gili Di Klu Menggunakan K-Means Clustering," *J. Teknol. Informasi, Komputer, Dan Apl.*, Vol. 4, No. 2, Pp. 177–189, 2025.

[16] A. Supriyadi, A. Triayudi, And I. D. Sholihati, "Perbandingan Algoritma K-Means Dengan K-Medoids Pada Pengelompokan Armada Kendaraan Truk Berdasarkan Produktivitas," *Jipi (Jurnal Ilm. Penelit. Dan Pembelajaran Inform.*, Vol. 6, No. 2, Pp. 229–240, 2021, Doi: 10.29100/Jipi.V6i2.2008.

[17] G. B. Kaligis And S. Yulianto, "Analisa Perbandingan Algoritma K-Means, K-Medoids, Dan X-Means Untuk Pengelompokkan Kinerja Pegawai," *It-Explore J. Penerapan Teknol. Inf. Dan Komun.*, Vol. 1, No. 3, Pp. 179–193, 2022, Doi: 10.24246/Itexplore.V1i3.2022.Pp179-193.

[18] K. C. Di, R. Sakit, W. Ngawi, H. Dilawati, H. Widianto, And A. Kuswiadji, "Klasterisasi Data Rekam Medis Pasien Menggunakan Metode K-Means Clustering Di Rumah Sakit Widodo Ngawi," *Teknol. Inf. Dan Rekayasa Komput.*, Vol. 5, No. 2, Pp. 139–147, 2024.