



Komparasi Metode Perhitungan Jarak K-Means Paling Baik Terhadap Pembentukan Pola Kunjungan Wisatawan Mancanegara

Lalu Mutawalli^{1,*}, Sofiansyah Fadli², Supardianto³

¹Prodi Sistem Informasi, STMIK Lombok, Praya

Jalan Basuki Rahmat Praya Mataram, Praya, Kec. Praya, Kabupaten Lombok Tengah, Nusa Tenggara Barat, Indonesia

²Prodi Teknik Informatika, STMIK Lombok, Praya

Jalan Basuki Rahmat Praya Mataram, Praya, Kec. Praya, Kabupaten Lombok Tengah, Nusa Tenggara Barat, Indonesia

³Prodi Teknik Informatika, Universitas Teknologi Mataram, Mataram, Indonesia

Jalan Basuki Rahmat Praya Mataram, Praya, Kec. Praya, Kabupaten Lombok Tengah, Nusa Tenggara Barat, Indonesia

Email: ^{1,*}laluallistilo@gmail.com, ²sofiansyah182@gmail.com, ³odyputra88@gmail.com,

Email Penulis Korespondensi: laluallistilo@gmail.com

Submitted: 07/10/2023; Accepted: 25/10/2023; Published: 26/10/2023

Abstrak-Mengetahui pola terhadap wisatawan manca Negara menjadi suatu hal yang urgent, pola tersebut dapat menjadi suatu pengetahuan yang dapat membantu untuk pengambilan keputusan yang lebih baik karena berbasis data-driven. Pola yang akan dielaborasi yaitu tentang pengelompokan kunjungan wisatawan mancanegara terhadap destinasi wisata di Jakarta. Data mining merupakan pendekatan yang membuat pola pengetahuan terhadap kumpulan data. K-Means merupakan salah satu algoritma data mining dalam melakukan pengelompokan (cluster) terhadap data, data akan dikelompokkan berdasarkan kesamaan fitur dan atributnya. Pada penelitian ini dilakukan perbandingan metode Euclidean Distance, Manhattan Distance, Haversine Distance tujuannya mendapatkan kelompok data yang lebih representatif terhadap datasets. Datasets pada penelitian ini memiliki tidak berdistribusi normal disebabkan oleh data outlier sehingga digunakan Algoritma DBSCAN perbaikan tanpa melakukan penghilangan dan pemotongan data karena dapat menyebabkan banyak nilai hilang yang berpengaruh terhadap informasi yang tidak sesuai dengan fakta empirik. Pada penelitian ini dibuat 5 cluster merujuk pada hasil perhitungan elbow, Berdasarkan hasil pengujian pada cluster K-Means pada Euclidean, mendapatkan Silhouette Score 0.36, Inertia 0.86, dan Davies-Bouldin Index 2.39, Metode Manhattan Silhouette Score 0.65, Inertia 1.46, dan Davies-Bouldin Index 0.47, sedangkan penerapan metode Haversine mendapatkan nilai Silhouette Score 0.36, Inertia 0.03, nilai 2.39 pada pengujian Davies-Bouldin Index.

Kata Kunci: K-Means; Cluster; Destinasi; Wisata; Mancanegara

Abstract-Understanding patterns among foreign tourists is an urgent matter. These patterns can become knowledge that helps in making better decisions because they are data-driven. The pattern to be elaborated on is regarding the clustering of visits by foreign tourists to tourist destinations in Jakarta. Data mining is an approach that extracts knowledge patterns from a dataset. K-Means is one of the data mining algorithms used for clustering data, where data is grouped based on similarity in features and attributes. This study compares the Euclidean Distance, Manhattan Distance, and Haversine Distance methods to obtain more representative data clusters for the datasets. The datasets in this study are not normally distributed due to outlier data; hence, the DBSCAN algorithm is used for improvement without removing or cutting the data, as it can result in a significant amount of missing values that could affect information that does not align with empirical facts. In this study, 5 clusters were created based on elbow calculation results. The K-Means cluster testing in Euclidean distance yielded a Silhouette Score of 0.36, Inertia of 0.86, and Davies-Bouldin Index of 2.39. The Manhattan method resulted in a Silhouette Score of 0.65, Inertia of 1.46, and Davies-Bouldin Index of 0.47. Meanwhile, applying the Haversine method resulted in a Silhouette Score of 0.36, Inertia of 0.03, and a value of 2.39 for the Davies-Bouldin Index.

Keywords: K-Means; Cluster; Destinations; Tourism; Foreign

1. PENDAHULUAN

Sektor pariwisata memiliki peran penting dalam pertumbuhan dan perkembangan pembangunan di Indonesia, pada bidang ekonomi pariwisata memiliki pengaruh positif dan signifikan[1]. Walau beberapa tahun tahun 2019-2021 menurun akibat pandemi di tahun 2023 pertumbuhannya mengalami peningkatan mencapai hingga 2,5 juta kunjungan[2]. Indonesia memiliki banyak destinasi yang indah berupa budaya, keindahan alam maupun destinasi wisata sejarah. Destinasi pariwisata merupakan kawasan geografis yang terletak di dalam satu wilayah maupun wilayah administratif lainnya yang memiliki daya Tarik[3]. Keberagaman destinasi wisata terdapat di setiap wilayah Republik Indonesia mulai dari perkotaan hingga pedesaan. Destinasi wisata kota Jakarta memiliki berbagai destinasi wisata yang memiliki daya Tarik terhadap wisatawan mancanegara. Berbagai kategori seperti tempat bersejarah, budaya, museum yang selalu menjadi alasan wisatawan untuk berwisata di ibukota Republik Indonesia.

Jakarta telah menerapkan open data melalui situs open data Jakarta[4] termasuk data tentang kunjungan pariwisata. Mengetahui pola terhadap wisatawan manca Negara menjadi suatu hal yang urgent, pola tersebut dapat menjadi suatu pengetahuan yang dapat membantu untuk pengambilan keputusan yang lebih baik karena berbasis data-driven. Pola yang akan dielaborasi yaitu tentang pengelompokan kunjungan wisatawan mancanegara terhadap destinasi wisata di Jakarta. Data mining merupakan pendekatan yang membuat pola pengetahuan terhadap kumpulan data. K-Means merupakan salah satu algoritma data mining dalam melakukan pengelompokan (cluster) terhadap data, data akan dikelompokkan berdasarkan kesamaan fitur dan atributnya[5]. Penelitian terkait dengan Penerapan K-Means dengan perhitungan jarak berbagai pendekatan perhitungan Antara lain Euclidean guna mengetahui tantangan pendidikan selama masa pandemi, khususnya dalam konteks penggunaan sistem online

learning Uhamka (OLU) berdasarkan hasil pengujian dengan indeks davies bouldin adalah 0.3806830859 dinyatakan cukup baik[6]. Melakukan improvisasi pada Euclidean K-Means dengan meningkatkan rasio aproksimasi 6.12903 memberikan kontribusi dalam memahami permasalahan algoritma K-Means[7].

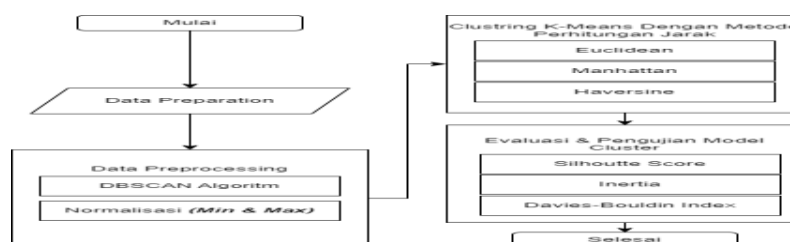
Pengelompokan data mengenai rumah sakit di Ibu Kota Jakarta, penegelompokan rumah sakit dengan menerapkan K-means dan metode Euclidean, rumah sakit dibagi menjadi tiga cluster dengan hasil evaluasi mean of square mendapatkan score 84,82%, 14,66%, dan 0.52% untuk cluster 0, 1, dan 2 secara berturut-turut[8]. Pada penelitian ini melakukan improvisasi pada algoritma K-means dengan metode Euclidean dan mendapatkan hasil pengujian dengan akurasi 80,57% pada Sembilan dataset mutivariat[9]. Perbandingan metode Euclidean dan Gower dengan menggunakan tujuh dataset berbeda yaitu data pelanggan grosir, data hepatitis, kontrasepsi, ujian nasional sekolah menengah pertama, penyakit diare, campak, dan data tenaga kesehatan, metode yang digunakan untuk pengujian adalah Silhouette pada metode Euclidean 0.8121 dan metode Gower mendapatkan 0.8513 pada dataset yang sama[10]. Penerapan K-Means dengan perhitungan jarak Manhattan memiliki kinerja yang baik dalam lingkungan big data[11]. Melakukan perbandingan Antara metode Euclidean dan Manhattan untuk melakukan pengelompokan data bunga irish, penelitian ini mengklaim bahwa metode Manhattan memiliki performa yang lebih baik[12]. Komprasi Manhattan, Euclidean, Minkoski dengan mempertimbangkan hasil evaluasi dengan metode Silhouette coefficient, ketiga metode memiliki hasil yang baik untuk data yang memiliki sedikit atribut sedangkan untuk data yang memiliki atribut yang banyak mendati nilai mendekati 0.5 cukup baik[13]. Perbandingan Antara metode Euclidean dan Manhattan untuk mendapatkan hasil cluster data pasien covid-19 di Kabupaten Malang, hasil pengujian menunjukkan Euclidean mendapatkan Silhouette Coefficient 0,71 Manhattan mendapatkan Silhouette Coefficient 0.64 sedangkan pada nilai standar deviasi Euclidean lebih kecil $0,46466002 < 0,4961977$ dengan hasil diperoleh metode Manhattan[14]. Pemantauan karyawan dengan melakukan cluster pada karywan yang seringkali bolos pada waktu kerja, dengan membandingkan tiga metode perhitungan jarak Euclidean. Manhattan dan Haversine untuk menemukan penentuan jarak yang paling baik, hasil pengujian metode Haversine unggul jika dibandingkan dengan metode Euclidean dan Manhattan dengan skor masing-masing Haversine 98,66%, Euclidian 83,33% sedangkan Manhattan 66,67% nilai diukur bedasarkan akurasi[15]. Melakukan analisa dengan memanfaatkan metode perhitungan Haversine pada jaringan Software Defined Networking (SDB) tujuannya untuk mengurangi latensi jaringan, selain itu dapat meingkatkan keandalan kontroler[16].

Pada penelitian yang telah dilakukan sebelumnya penerapan metode perhitungan jarak hanya digunakan secara parsial, tanpa melakukan komprasi, sehingga belum didapatkan justifikasi metode perhitungan jarak yang paling ideal dalam membuat cluster. Meninjau dari aspek ini perlu untuk melakukan suatu uji coba terhadap bebrapa metode perhitungan jarak untuk memastikan metode perhitungan jarak yang paling ideal dalam melakukan pengelompokan data khususnya pada data destinasi wisata. Sehingga Pada penelitian ini dilakukan dengan motivasi utama membandingkan metode Euclidean Distance, Manhattan Distance, Haversine Distance tujuannya mendapatkan kelompok data yang lebih representatif terhadap datasets. pengelompokan yang lebih baik dapat memberikan pengetahuan tentang pola kunjungan wisatawan mancanegara terhadap destinasi wisata Jakarta. Pengelompokan data yang akurat dan tepat dapat memberikan pengetahuan tentang pola dan karakteristik wisatawan mancanegara khususnya pada kecendrungan destinasi wisata, sehingga memberikan insight pengetahuan yang dapat digunakan untuk bahan pertimbangan peprbaiki keputusan dan pengelolaan bidang pariwisata khususnya di Ibu Kota Jakarta.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Tahapan pada penelitian mencakup enam aspek, tahap pertama melakukan pengumpulan data, kedua melakukan data preparation dan preprocessing data dengan menerapkan metode DBSCAN, ketiga menentukan jumlah K yang ideal menggunakan metode elbow, keempat membuat cluster dengan K-Means , kelima melakukan clustering dengan metode perhitungan jarak Antara lain metode Euclidean Distance, Manhattan Distance, Haversine Distance, kelima melakukan evaluasi dan komparasi terhadap clustering yang dibuat menggunakan ketiga metode perhitungan Euclidean Distance, Manhattan Distance, Haversine Distance untuk menemukan hasil cluster yang paling baik. Gambar 1 merupakan tahapan penelitian.



Gambar 1. Tahapan Penelitian



2.2 Data Preparation

Data yang digunakan pada penelitian ini adalah data kunjungan wisatawan mancanegara terhadap 20 destinasi wisata yang ada di Jakarta. Sumber data berasal dari website Jakarta open data tahun 2020[17] dan data kunjungan wisatawan tahun 2021[18]. Terdapat 20 Destinasi wisata dan kunjungan wisatawan mancanegara sebanyak 46.165 orang. Table 1. menunjukkan data kunjungan wisatawan mancanegara pada tahun 2020-2021.

Tabel 1. Dataset yang digunakan Data Wisatawan Mancanegara 2020-2021

No	Destinasi Wisata	Jumlah Wisatawan Mancanegara
1	Kawasan Kota Tua	18621
2	Kepulauan Seribu	5211
3	Monumen Nasional	4339
4	Museum Bahari	640
5	Museum Joang	33
6	Museum Nasional	5974
7	Museum Prasasti	80
8	Museum Satria Mandala	0
9	Museum Sejarah Jakarta	4017
10	Museum Seni Rupa dan Keramik	866
11	Museum Tekstil	579
12	Museum Wayang	3026
13	PBB Setu Babakan	12
14	Pelabuhan Sunda Kelapa	180
15	Planetarium	0
16	Rumah si pitung	1
17	Taman Arkeologi Onrust	9
18	Taman Impian Jaya Ancol	1076
19	Taman Margasatwa Ragunan	1501
20	Taman Mini Indonesia Indah	0

2.3 Data Preprocessing

Pada tahap ini dilakukan proses data preprocessing karena terdapat data outlier, data outlier akan berpengaruh terhadap kualitas model menjadi tidak baik[19]. Perbaikan terhadap data outlier menjadi penting untuk dilakukan. Algoritma Density-Based Spatial Clustering of Application with Noise (DBSCAN) dapat digunakan untuk mengidentifikasi data outlier, tidak hanya memiliki kemampuan dalam mengidentifikasi juga dapat digunakan untuk memperbaiki data outlier pada sebuah datasets[20].

2.4 Menentukan Nilai K Dengan Elbow

Pemilihan K atau jumlah cluster dengan cara manual atau mengira-ngira akan terkesan subjektif, pada penelitian ini pemilihan K didasarkan kepada analisa menggunakan metode “Elbow”, tahapan penentuan jumlah K berdasarkan metode elbow Antara lain[5]:

- Preparation data, data sebelumnya telah diolah menggunakan algoritma DBSCAN sehingga lebih layak untuk di proses ketahap penentuan K;
- Mentukan rentang K dan melakukan perhitungan inertia dengan persamaan.

$$CSS(K) = \sum_{i=1}^K \sum_{j=1}^{n_i} \|x_{ij} - \mu_i\|^2 \tag{1}$$

- Buat visualisasi dengan style plot dari nilai K terhadap WCSS akan menghasilkan suatu kurva yang kemudian disebut sebagai elbow curve.
- Analisis kurva dan penentuan nilai K – Optimal yang telah dibuat untuk menjustifikasi jumlah K yang ideal, biasanya tergambar dengan titik menurun dan tampak landai.
- Implementasi K-Means dengan tiga opsi perhitungan jarak pada kasus penelitian ini adalah Euclidian Distance, Manhattan Distance, Haversine Distance.

2.5 Pembuatan Cluster K-Means

Pada penelitian ini pengelompokan data menggunakan algoritma K-Means dengan pendekatan tiga jenis metode perhitungan jarak. Tahapan pengelompokan menggunakan K-Means Antara lain[21]:

- Inisialisasi pusat cluster;
- Perhitungan jarak, pada tahap ini menggunakan Euclidian Distance, Manhattan Distance, Haversine Distance;
- Perbahruai pusat cluster;
- Melakukan perulangan pada tahap 2 perhitungan jarak dan 3 pemusatan data;
- Evaluasi dan menentukan hasil akhir.



2.5.1 K-Means dengan Euclidian Distance

Euclidian merupakan konsep geometri untuk mengukur jarak, pada runag (2D), P1 = (x1 - y1) dan P2 = (x2 - y2) pada ruang dua dimensi akar kuadrat x dan y diantara kedua titik, persamaan adalah sebagai berikut:

Euclidean = sqrt((x2 - x1)^2 + (y2 - y1)^2) (2)

Sedangkan pad ruang (3D), P1 = (x1, y1, z1) dan P2 = (x2, y2, z2) pada ruang dua dimensi akar kuadrat x dan y diantara kedua titik, persamaan adalah sebagai berikut[21]:

Euclidean = sqrt((x2 - x1)^2 + (y2 - y1)^2 + (z2 - z1)^2) (3)

2.5.2 K-Means dengan Manhattan Distance

Manhattan distance familiar juga dikenal dengan istilah city block merupakan jarak alternatif digunakan dalam berbagai konteks yang bertujuan untuk analisis dan pemodelan. Rumus dasar manhattan adalah sebagai berikut:

Manhattan = sum_{i=1}^n |x2i - x1i| (4)

2.5.3 K-Means dengan Haversine Distance

Haversine distance merupakan metodologi perhitungan jarak mempertimbangkan dua titik dipermukaan bola katakanlah (bumi) menggunakan koordinat lintang dan bujur. Pada haversine R merupakan jari-jari bola sedangkan phi1, phi2, lambda1 dan lambda2 merupakan kordinat lintang dan bujur dalam radian dari dua titik yang ingin dilakukan pengukuran jaraknya. Rumus haversine menggunakan pendekatan perhitungan melalui suatu lingkaran besar pada suatu permukaan bola, hasil dari peritungan jarak dapat dinyatakan dengan satuan kilometer atau juga mil. Untuk menkonversi derajat dari redian digunakan rumus sebagai berikut:

Radian = (Derajat) x (pi/180) (5)

2.6 Metode Evaluasi

Evaluasi terhadap model clustering dengan algoritma K-Means mutlak dilakukan untuk mendapatkan sejauh mana model cluster terbentuk dengan hasil yang baik. Pada penelitian ini membangun cluster kunjungan wisatawan manca Negara berdasarkan destinasi wisata Ibu Kota Jakarta. Penerapan tiga metode perhitungan jarak dilakukan untuk menemukan metode yang paling tepat pada kasus ini, cluster yang telah dibuat yang didasari tiga metode kemudian dievaluasi dengan menggunakan tiga metode pengujian yaitu, Silhouette Score, Inertia Davies-Bouldin Index.

a. Silhouette Score

Silhouette Score merupakan salah satu metode pengujian model pada machine learning, tujuannya mengevaluasi sejauh mana objek berada pada suatu cluster. Terdapat dua cara dalam melakukan pengukuran yaitu 1) Rata-rata jarak Antara sebuah sampel dengan semua sampel pada cluster; 2) Rata-rata jarak antara sebuah sampel dengan klaster paling dekat. Pada rumus 3.5 adalah persamaan metode Slhouete Score[22].

S = (b-a) / max(a,b) (6)

b. Inertia

Metode inertia adalah merupakan salah satu metode yang digunakan untuk mengukur sejauh mana titik-titik data di dalam sebuah cluster. Inertia membantu untuk melakukan evaluasi kompaknes dari suatu cluster. Kompaknes adalah keeratan titik-titik data yang terkumpul disekitar centroid, rumus 3.6 adalah merupakan persamaan umum untuk menghitung inertia (I) pada cluster[20].

I = sum_{i=1}^K sum_{x in C1} ||x - mu_i||^2 (7)

c. Davies-Bouldin Index

Davies-Bouldin Index atau biasa disingkat DBI adalah matrix untuk melakukan evaluasi clustering dengan konsep seberapa baik pembagian data kedalam kelompok tertentu, DBI yang lebih rendah mengindikasikan pengelompokan data yang baik. Rumus DBI 3.7 merupakan persamaan Davies-Bouldin Index[5].

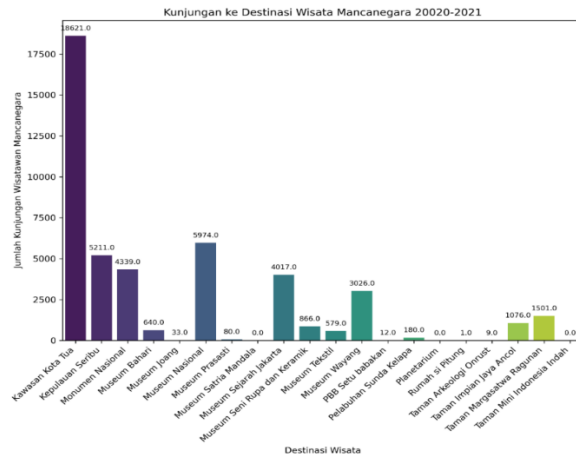
DBI_i = R_i / (1/C1 * sum_{j=1}^K d(c_i, c_j)) (8)

3. HASIL DAN PEMBAHASAN

3.1 Ekplorasi Dataset

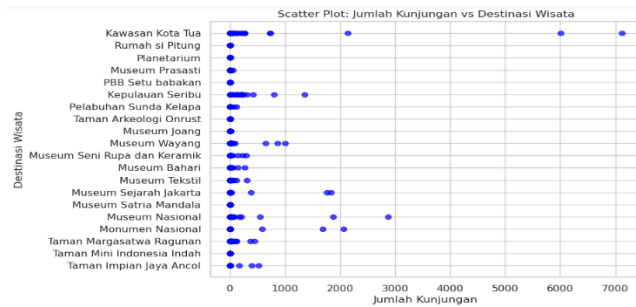
Jumlah data yang terdapat pada dataset sebanyak 24 field, data yang digunakan selama 24 bulan terhitung Januari 2020 sampai dengan Desember 2021. Berdasarkan dataset 24 bulan jumlah wisatawan yang berkunjung ke 20

destinasi wisata Jakarta sebanyak 46.165 wisatawan mancanegara. Berdasarkan grafik pada gambar 4.1 masing-masing wisata mendapat kunjungan sebagai berikut Kawasan Kota Tuan 18.621, Kepulauan Seribu 5.211, Monumen Nasional 4.339, Museum Bahari 640, Museum Joang 33, Museum Nasional 5.974, Museum Prasasti 80, Museum Satria Mandala 0, Museum Sejarah Jakarta 4.017, Museum Seni Rupa dan Keramik 866, Museum Tekstil 579, Museum Wayang 3.026, PBB Setu Babakan 12, Pelabuhan Sunda Kelapa 180, Planetarium 0, Rumah si Pitung 1, Taman Arkeologi Onrust 9, Taman Impian Jaya Ancol 1.076, Taman Marga Satwa Ragunan 1.501, Taman Mini Indonesia Indah 0 Kunjungan.



Gambar 2. Grafik Kunjungan Wisatawan Mancanegara

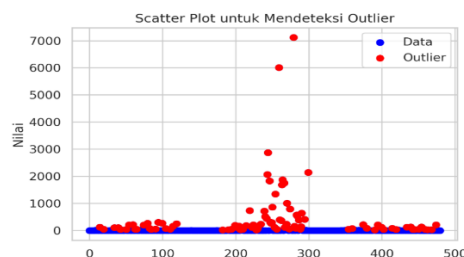
Persebaran data dilakukan untuk mengetahui distribusi pada dataset kunjungan wisatwan manca Negara terhadap destinasi wisata Jakarta. Pada tahap analisis persebaran data dilakukan visualiasi dengan scatter plot dengan tujuan menemukan distribusi data. Berdasarkan gambar 2 dapat dilihat bahwa data tidak berdistribusi normal, tidak berdistribusi normal dikarenakan terdapat suatu destinasi yang sangat banyak dikunjungi seperti Kawasan Kota Tua memiliki nilai paling tinggi melampaui data yang lainnya, Terdapat destinasi wisata yang nilai kunjungannya nol. Berdasarkan permasalahan distribusi data diperlukan tindakan lebih lanjut untuk memperbaiki data agar dapat dilakukan analisa atau pembuatan model cluster. Upaya yang dilakukan pada penelitian ini dengan menerapkan algoritma DBSCAN untuk melakukan perbaikan terhadap datasets.



Gambar 3. Distribusi persebaran data

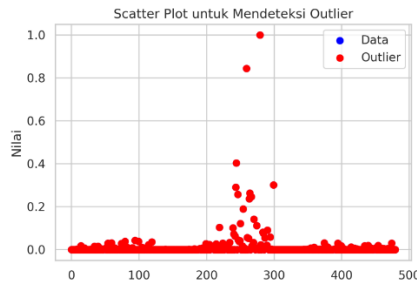
3.2 Data Preprocessing

Berdasarkan hasil analisa pada tahap sebelumnya ditemukan bahwa data tidak berdistribusi normal dikarenakan terdapat data outlier. Analisa lebih lanjut tentang data outlier dilakukan untuk mengetahui fitur-fitur yang memiliki nilai outlier. Pada gambar 4 merupakan hasil visulisasi terhadap data dengan melakukan visualisasi menggunakan scatter plot. Data yang normal diinisialisasi sebagai lingkaran berwarna biru sedangkan data outlier diinisialisasi dengan lingkaran berwarna merah.



Gambar 4. Pendeteksian Data Outlier

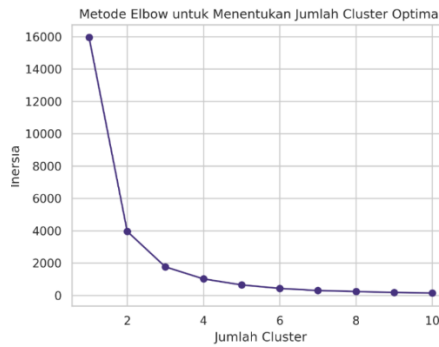
Berdasarkan hasil visualisasi pada gambar 4 ditemukan data outlier yang cukup banyak, data outlier akan berdampak tiak baik apabila dilanjutkan pada tahap pembuatan model cluster. Penerapan Algoritma DBSCAN merupakan upaya yang dapat dilakukan sehingga data harus dilakukan perbaikan tanpa melakukan penghilangan dan pemotongan data karena dapat menyebabkan banyak nilai hilang yang berpengaruh terhadap informasi yang tidak sesuai dengan fakta empirik. Penerapan algoritma DBSCAN merupakan alternative sebagai solusi yang baik pada kaus ini, hasil perbaikan dengan menerapkan algoritma DBSCAN pada data pada dataset ditunjukkan pada gambar 5 dengan asumsi bahwa melakukan pengisian data pada data missing dengan mempertimbangkan centroid terdekatnya, sehingga data yang hilang akan direpresentasikan dengan karakteristik yang lebih dekat dengan nilai sekitarnya.



Gambar 5. Hasil perbaikan Pada Data outlier

3.3 Model Cluster

Pembuatan jumlah kelompok (K) dengan metode manual seringkali menimbulkan asumsi subjektif karena sifatnya dilakukan dengan cara perkiraan, pada penelitian ini untuk penentuan K digunakan metode Elbow agar proses lebih objektif karena berlandaskan pada hasil perhitungan yang definitif. Berdasarkan gambar 6 dengan mempertimbangkan justifikasi metode elbow jumlah K yang ideal ditandai dengan pola grafik melandai tidak curam. Nilai K yang ideal menurut hasil analisa menggunakan metode elbow adalah 5 cluster, pada grafik menunjukkan Antara 1 hingga 4 masih menunjukkan grafik masih curam, ketika grafik terletak pada nilai 4 grafik masih belum melandai dan pada nilai 5 tampak melandai sehingga dapat diratik justifikasi nilai 5 merupakan nilai yang stabil dan yang paling ideal dalam pembuatan jumlah kelompok.



Gambar 6. Jumlah K yang ideal

3.4 Hasil Clustring

Pada tahap sebelumnya telah dilakukan analisa menggunakan metode elbow sehingga didapati jumlah cluster yang akan dibuat pada penelitian ini sebanyak 5 cluster. Pada tahap selanjutnya cluster dibangun dengan menerapkan algoritma K-Means, pada tahap perhitungan jarak di algoritma K-Means dilakukan simulasi dengan menerapkan metode Euclidean, Manhttan, dan Haversine secara bergantian. Gambar 7 merupakan visualisasi hasil cluster dengan algoritma K-Means.



Gambar 7. Visualisasi Cluster

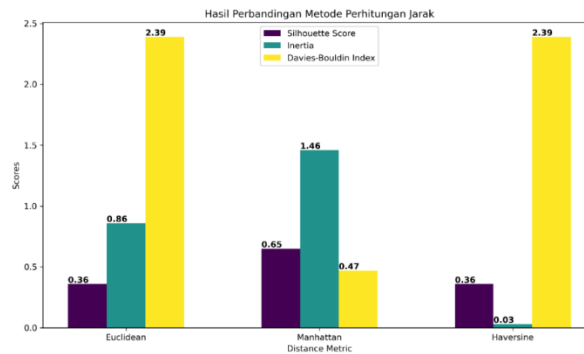
Hasil cluster menggunakan K-Mean menjadi enam cluster, pada cluster 1 terdapat 6 destinasi wisata, cluster 2 terdapat 6 destinasi wisata, cluster 3 terdapat 1 destinasi wisata, cluster 4 terdapat 1 destinasi wisata, dan cluster 5 terdapat 6 destinasi wisata. Pada table 4.1 dapat dilihat lebih rinci masing-masing destinasi wisata pada setiap cluster.

Tabel 2. Menunjukkan hasil cluster destinasi wisata

Jumlah Destinasi	Nilai Cluster	Nomor Cluster	Destinasi_Wisata
6	0.108992	1	Monumen Nasional, Museum Nasional, Museum Satria Mandala, Museum Bahari, Museum Joang, Museum Prasasti
6	0.16485	2	Taman Impian Jaya Ancol, Taman Mini Indonesia Indah, Taman Margasatwa Ragunan, Taman Arkeologi Onrust, Planetarium, Rumah si Pitung
1	1	3	Kawasan Kota Tua
1	0.414169	4	Kepulauan Seribu
6	0.133515	5	Museum Sejarah Jakarta, Museum Tekstil, Museum Seni Rupa dan Keramik, Museum Wayang, Pelabuhan Sunda Kelapa, PBB Setu babakan

3.5 Hasil Evaluasi

Cluster yang telah dibangun menggunakan K-Means dengan menerapkan Euclidean, Manhattan, dan Haversine dievaluasi menggunakan tiga metode evaluasi yaitu metode Silhouette Score, Inertia, dan Davies-Boldin Index. Pada gambar 4.7 merupakan nilai hasil evaluasi setiap metode. Berdasarkan hasil pengujian pada cluster K-Means dengan menerapkan metode perhitungan jarak Euclidean source code pembuatan model dan pengujian dapat diakses pada link <https://github.com/alimahatma/komparasi-metode-perhitunganjarak/blob/main/KMeans-Euclidean.ipynb> mendapatkan Silhouette Score 0.36, Inertia 0.86, dan Davies-Bouldin Index 2.39. Pada penerapan metode Manhattan mendapatkan nilai Silhouette Score 0.65, Inertia 1.46, dan Davies-Bouldin Index 0.47 proses pembuatan model dan pengujian dapat diakses melalui link berikut <https://github.com/alimahatma/komparasi-metode-perhitunganjarak/blob/main/KMeans-Manhattan.ipynb>, sedangkan penerapan metode Haversine mendapatkan nilai Silhouette Score 0.36, Inertia 0.03, dan Davies-Bouldin Index 2.39 proses pengujian dapat diakses link <https://github.com/alimahatma/komparasi-metode-perhitunganjarak/blob/main/KMeans-Haversine.ipynb>.



Gambar 8. Hasil Evaluasi

4. KESIMPULAN

Pada penelitian ini dibuat lima cluster yang merujuk pada hasil analisa menggunakan metode elbow dengan masing-masing cluster memiliki anggota Antara lain, cluster 1 Monumen Nasional, Museum Nasional, Museum Satria Mandala, Museum Bahari, Museum Joang, Museum Prasasti, cluster 2 memiliki anggota kelompok Taman Impian Jaya Ancol, Taman Mini Indonesia Indah, Taman Margasatwa Ragunan, Taman Arkeologi Onrust, Planetarium, Rumah si Pitung, cluster 3 hanya satu destinasi wisata Kawasan Kota Tua, cluster 4 satu destinasi wisata Kepulauan Seribu, dan cluster 5 terdapat enam anggota kelompok Museum Sejarah Jakarta, Museum Tekstil, Museum Seni Rupa dan Keramik, Museum Wayang, Pelabuhan Sunda Kelapa, PBB Setu babakan. Pada penelitian ini melakukan komparasi terhadap tiga metode perhitungan jarak. Hasil komparasi terhadap ketiga metode tersebut didapatkan hasil metode Euclidean menjadi metode yang paling ideal karena memiliki nilai yang baik pada tiga metode evaluasi. Haversine menjadi metode perhitungan jarak yang kedua dikarenakan memiliki score yang sama pada Silhouette Score dan Davies-Bouldin Index yang sama dengan hasil Euclidean, dan metode



Manhattan menjadi metode yang memiliki score paling rendah pada Shilhouette Score dan Davies-Bouldin Index namun memiliki nilai paling tinggi pada pengujian Inertia.

REFERENCES

- [1] D. Rahmayani, S. Oktavilia2, D. A. Suseno, E. L. Isnaini, and A. Supriyadi, “Economics Development Analysis Journal Tourism Development and Economic Growth: An Empirical Investigation for Indonesia Article Information,” *Econ. Dev. Anal. J.*, vol. 1, no. 1, pp. 1–11, 2022, [Online]. Available: <https://doi.org/10.15294/edaj.v1i1i1.50009>
- [2] P. Widayanti, “Kian Melesat di 2023 Wisata Indonesia Bersiap Menuju Level Pandemi,” *Media Keuangan Kemenku RI*, 2023. <https://mediakeuangan.kemenkeu.go.id/article/show/kian-melesat-di-2023-pariwisata-indonesia-bersiap-menuju-level-prapandemi>
- [3] Kemenkumham, “Undang-Undang Republik Indonesia Nomor 10 Tahun 2009 Tentang Kepariwisata,” *Kementerian Keuangan RI*, 2009. <https://jdih.kemenkeu.go.id/fullText/2009/10TAHUN2009UU.HTM#:~:text=Daerah tujuan pariwisata yang selanjutnya,saling terkait dan melengkapi terwujudnya>
- [4] DKI, “Data Terbuka Pemerintah Provinsi DKI Jakarta,” *Provinsi Jakarta*, 2020. <https://data.jakarta.go.id/>
- [5] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York: Springer-Verlag, 2006.
- [6] L. Ardiansyah and S. A. Awalludin, “Implementation of the K-Mean Algorithm to Determine the Level of Student Satisfaction with the Online Learning Uhamka System (OLU),” *J. Pembelajaran Dan Mat. Sigma*, vol. 9, no. 1, pp. 162–171, 2023, doi: 10.36987/jpms.v9i1.4121.
- [7] F. Grandoni, R. Ostrovsky, Y. Rabani, L. J. Schulman, and R. Venkat, “A refined approximation for Euclidean k-means,” *Inf. Process. Lett.*, vol. 176, p. 106251, 2022, doi: 10.1016/j.ipl.2022.106251.
- [8] K. E. Setiawan, A. Kurniawan, A. Chowanda, and D. Suhartono, “Clustering models for hospitals in Jakarta using fuzzy c-means and k-means,” *Procedia Comput. Sci.*, vol. 216, no. 2022, pp. 356–363, 2023, doi: 10.1016/j.procs.2022.12.146.
- [9] N. H. M. M. Shrifan, M. F. Akbar, and N. A. M. Isa, “An adaptive outlier removal aided k-means clustering algorithm,” *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 8, pp. 6365–6376, 2022, doi: 10.1016/j.jksuci.2021.07.003.
- [10] A. Aditya, N. B. Sari, and T. N. Padilah, “Perbandingan pengukuran jarak Euclidean dan Gower pada kluster k-medoids,” *J. Teknol. dan Sist. Komput.*, vol. 9, no. 1, pp. 1–7, 2021, doi: 10.14710/jtsiskom.2021.13747.
- [11] T. M. Ghazal et al., “Performances of k-means clustering algorithm with different distance metrics,” *Intell. Autom. Soft Comput.*, vol. 30, no. 2, pp. 735–742, 2021, doi: 10.32604/iasc.2021.019067.
- [12] R. Suwanda, Z. Syahputra, and E. M. Zamzami, “Analysis of Euclidean Distance and Manhattan Distance in the K-Means Algorithm for Variations Number of Centroid K,” *J. Phys. Conf. Ser.*, vol. 1566, no. 1, 2020, doi: 10.1088/1742-6596/1566/1/012058.
- [13] R. Hidayati, A. Zubair, A. H. Pratama, and L. Indana, “Analisis Silhouette Coefficient pada 6 Perhitungan Jarak K-Means Clustering Silhouette Coefficient Analysis in 6 Measuring Distances of K-Means Clustering,” *Techno.COM*, vol. 20, no. 2, pp. 186–197, 2021.
- [14] W. Wahyu Pribadi, A. Yunus, and A. S. Wiguna, “Perbandingan Metode K-Means Euclidean Distance Dan Manhattan Distance Pada Penentuan Zonasi Covid-19 Di Kabupaten Malang,” *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 6, no. 2, pp. 493–500, 2022, doi: 10.36040/jati.v6i2.4808.
- [15] Y. Miftahuddin, S. Umaroh, and F. R. Karim, “Perbandingan Metode Perhitungan Jarak Pada Kehadiran Karyawan Institut Teknologi Nasional Bandung,” *J. Tekno Insentif*, vol. 14, no. 2, pp. 69–77, 2020.
- [16] V. S. Thalapala and K. Guravaiah, “FCMCP: Fuzzy C-Means for Controller Placement in Software Defined Networking,” *Procedia Comput. Sci.*, vol. 201, no. 1, pp. 109–116, 2022, doi: 10.1016/j.procs.2022.03.017.
- [17] P. D. Jakrata, “Data Jumlah Kunjungan Wisatawan Mancanegara ke Destinasi Wisata di Provinsi DKI Jakarta Tahun 2020,” 2020. <https://data.jakarta.go.id/dataset/data-jumlah-kunjungan-wisatawan-mancanegara-ke-destinasi-wisata-di-provinsi-dki-jakarta-tahun-2021>
- [18] P. D. Jakarta, “Data Jumlah Kunjungan Wisatawan Mancanegara ke Destinasi Wisata di Provinsi DKI Jakarta Tahun 2021,” *Jakarta Open Data*, 2021. <https://data.jakarta.go.id/dataset/data-jumlah-kunjungan-wisatawan-mancanegara-ke-destinasi-wisata-di-provinsi-dki-jakarta-tahun-2021>
- [19] F. Jin, M. Chen, W. Zhang, Y. Yuan, and S. Wang, “Intrusion detection on internet of vehicles via combining log-ratio oversampling, outlier detection and metric learning,” *Inf. Sci. (Ny.)*, vol. 579, pp. 814–831, 2021.
- [20] P.-N. Tan, M. Steinbach, and V. Kumar, *Introduction to Data Mining*. 2005.
- [21] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. London: MIT Press, 2012.
- [22] C. Aggarwal, *Data Clustering Algorithms and Applications*. Florida: CRC Press, 2014.