

Implementasi Data Mining Untuk Pengelompokan Kunjungan Wisata Di Kabupaten Mojokerto

Mohamad Adi Putra Firdaus¹, Yuni Yamasari²

^{1,2} S1 Teknik Informatika, Fakultas Teknik, Universitas Negeri Surabaya

¹mohamad.20014@mhs.unesa.ac.id

²yuniyamasari@unesa.ac.id

Abstrak— Dinas Kebudayaan, Olahraga dan Pariwisata kabupaten Mojokerto menghadapi tantangan besar dalam mengelola dan mengolah data yang sangat banyak terkait tempat-tempat wisata di wilayahnya. Pengelolaan data yang kurang efektif dapat menghambat pengembangan strategi pariwisata yang komprehensif dan pemerataan kunjungan wisatawan di seluruh destinasi. Permasalahan ini membutuhkan solusi yang dapat mengoptimalkan proses pengolahan data sehingga dapat digunakan untuk merumuskan kebijakan. Selain itu, langkah pengembangan pariwisata yang tepat juga diperlukan untuk meningkatkan daya tarik destinasi wisata, dan memastikan distribusi kunjungan yang lebih merata di seluruh tempat wisata di kabupaten Mojokerto. Untuk itu, penelitian ini memfokuskan pada pengelompokan kunjungan wisata dengan K-Means. Proses pengoptimalan dilakukan dengan metode elbow untuk menentukan banyaknya klaster yang terbaik. Selanjutnya, analisis pengklasteran dilakukan dengan indek silhouette. Hasil penelitian dengan 14 kali uji coba memperlihatkan bahwa banyaknya klaster terbaik adalah 2 klaster baik dengan inisialisasi random dan inisialisasi kmeans++. Hal ini ditunjukkan dengan nilai silhouette tertinggi sebesar 0.6817446298183374 dengan hasil klaster 1 sebanyak 37 tempat wisata antara lain wisata parimas waterpark, candi tikus, candi brahu dan seterusnya. Sedangkan, klaster 2 sebanyak 3 tempat wisata, yaitu: pemandian air panas, wana wisata padusan dan makam troloyo.

Kata Kunci—wisata, K-Means, Data Mining, Elbow, silhouette.

I. PENDAHULUAN

Mojokerto adalah tujuan wilayah tempat wisata Indonesia dan tujuan wisata terpopuler dari dan Jumlah kunjungan wisatawan meningkat setiap tahun, menunjukkan daya tarik yang Semakin bertambah seiring berjalannya waktu. bisa dikatakan Mojokerto merupakan salah satu pondasi wisata di Indonesia yang dimana salah satu wisatanya masuk nominasi wisata dengan desa terbaik wisata tingkat nasional. Kabupaten Mojokerto mempunyai segudang keindahan, serta kekayaan wisata alam dan sejarah. Potensi mendasar ini bisa di dapat dan dimanfaatkan sebagai destinasi wisata dan sumber pendapatan daerah demi menunjang keberlanjutan wisata yang ada[1]. Kemajuan maupun perkembangan dalam sektor pariwisata menjadi salah satu akar atau bagian dalam membangun pondasi perekonomian dalam suatu daerah Pembangunan pariwisata sangatlah penting [2].

Dengan meningkatnya minat wisata di Kabupaten Mojokerto, diperlukan suatu sistem yang dapat membantu mengolah juga mengelompokkan kunjungan wisatawan berdasarkan karakteristik tertentu untuk mengetahui destinasi wisata mana yang tergolong destinasi wisata dengan jumlah pengunjung tinggi dan pengunjung rendah. hal tersebut dapat

dilakukan agar setiap wisata berlomba - lomba dalam bagaimana agar wisatanya dapat memiliki daya tarik yang tinggi [3]. Dalam pengelompokan terdapat suatu proses data mining dimana melibatkan penambangan atau penemuan informasi baru. data mining, sendiri merupakan suatu alat atau aplikasi yang akan digunakan analisis statistik atau kode pada data [4]. data mining bukanlah suatu topik baru dalam penelitian, data mining sendiri biasanya digunakan untuk meningkatkan akurasi [5]. dalam rangkaian proses data mining, terdapat beberapa jenis metode seperti estimasi, prediksi, klasifikasi, pengelompokan, dan asosiasi [6].

Dalam penelitian ini Metode yang dipakai untuk mengategorikan destinasi pariwisata adalah menggunakan K-means clustering. algoritma K-Means mengevaluasi setiap bagian dalam set data dan menggolongkannya ke dalam pusat kluster yang telah ditentukan berdasarkan jarak minimum antara bagian tersebut dan setiap pusat kluster [7]. teknik yang digunakan dalam alur algoritma K-Means dapat berupa mengelompokkan data ke dalam kelompok-kelompok yang masing-masing kelompok mempunyai karakteristik yang sama, namun ada pula data yang memiliki karakteristik yang berbeda dengan kelompok lainnya [8]. dalam kmeans dua objek atau titik data dikatakan serupa jika jaraknya cukup dekat satu sama lain [9] artinya dapat diartikan, semakin dekat jarak antar dua objek atau lebih pada benda, maka semakin mirip kedua objek atau benda tersebut. Semakin besar nilai jarak maka semakin besar ketidaksamaannya. algoritma K-means menugaskan setiap elemen ke cluster dengan centroid (mean) terdekat. Sederhananya, inilah langkah-langkah dalam proses pengelompokan K-means [10].

Dalam K-Means dalam hal memperoleh jumlah nilai k, kita bisa memperolehnya dengan sebuah metod yang sangat fleksibel yang dikenal sebagai elbow. metode elbow merupakan proses yang diperuntukan untuk memprediksi jumlah kelompok yang optimal dalam algoritma k-means clustering [11]. metode ini menghasilkan sebuah informasi atau data dalam menentukan berapa jumlah cluster cluster terbaik yang dapat dilihat dengan cara melihat persentase dari hasil perbandingan sehingga kita dapat mencerna atau mengetahui cluster mana yang memiliki kriteria yang ingin dikembangkan atau di Analisa [12]. Setelah itu di perlukan proses menghitung siluet setiap cluster atau keseluruhan cluster berdasarkan kerja algoritma clustering [13].

Dalam uraian mengenai latar belakang di atas maka didapat, Implementasi data mining dapat membantu dalam membuat prediksi terkait tren kunjungan wisata karena ketidak merataannya kunjungan wisata dan pengembangan kebijakan pariwisata di Kabupaten Mojokerto . dalam Penelitian ini

peneliti juga berharap hasil hail clustering dengan menggunakan metode alghortima K-Means dengan optimasi jumlah jumlah cluster dengan menggunakan proses metoden elbow dapat menghasilkan sebuah data berbentuk cluster yang optimum [14].

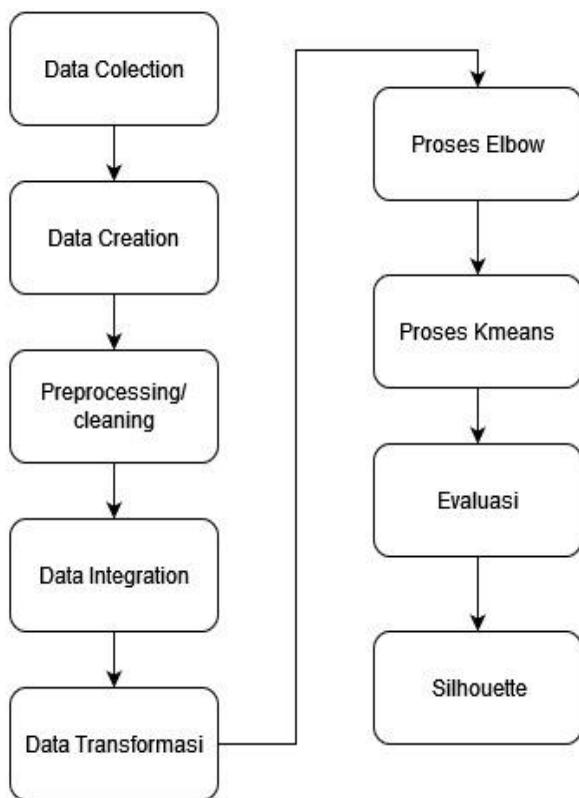
II. MATERI DAN METODE

A. Pengumpulan Data

Dalam proses ini, langkah yang dilakukan penulis yaitu dengan melaksanakan pengambilan data kepada berbagai pihak tertentu yang memiliki kaitan untuk pengambilan data tersebut. dalam proses penelitian ini sebagai upaya untuk memperoleh informasi terkait dengan data data yang akan digunakan oleh peneliti pada tahap selanjutnya. dimana pihak adalah dinas bagian pariwisata kabupaten mojokerto.

B. Metode Yang Dihasilkan

Dalam penelitian ini adalah mengenai implementasi data mining pada wisata yang ada di mojokerto dimana peneliti akan melakukan perancangan dan pengimplementasian alg/////////oritma machine learning pada dataset yang bertujuan untuk mengelompokkan wisata di Mojokerto menggunan k-means clustering. untuk proses penelitian dapat di lihat pada gambar 1.



Gbr. 1 Alur penelitian

alur gambar 1 untuk metode penelitian yang di lakukan oleh peneliti adalah KDD (Knowledge Discovery In Database) dimana untuk langkah langkah penelitian nya adalah :

1) Data Collection

Dalam proses data collection peneliti mengambil lagi data ke dinas pariwisata kabupaten Mojokerto karena di rasa dalam pengambilan data sebelumnya belum memenuhi kriteria dalam penelitian setelah itu di dapat data yang terkumpul yaitu data kunjungan wisata tahun 2018 -2023.

2) Data creation

Dan di tahap ini, peneliti memilih, mengambil dan menggunakan data yang sejenis dan sesuai dengan objek penelitian yang di teliti oleh peneliti. di tahap tahap ini penulis mengambil dan menghimpun data-data yang terdapat pada dinas bagian pariwisata kabupaten mojokerto. Data yang diambil berupa data pengunjung pariwisata tahun 2018-2023 yang terdiri dari pengunjung wisatawan mancanegara dan wisatawan nusantara yang bakalan akan dipakai dalam proses penerapan data mining.

2) Preprocessing/Cleaning

Pada tahap preprocessing/cleaning dilakukanlah alur alur dalam hal proses untuk membuat dataset yang memiliki nilai bersih untuk digunakan pada tahap selanjutnya yaitu tahap alur yang bertujuan untuk mendapatkan informasi yang berguna. dimana dalam tahap ini mencari data yang di anggap tidak ada atau null, data ganda dan lain sebagainya pada dataset.

3) Data Integration

Pada alur data integrasi, dilakukan peroses data dari berbagai sumber data atau platform yang berbeda untuk membentuk satu set data yang terpadu. data ini telah melalui proses sebelumnya dan akan dipakai dalam mengolah data dengan menerapkan metode K-Means clustering.

4) Data Transformasi

Pada tahap transformasi adalah alur mengubah atau memanipulasi data dari satu bentuk atau format ke bentuk atau format lainnya sebelum di olah. dimana dalam penelitian ini dataset yang berformat (.xlsx) akan diubah kembali menjadi format "csv" (comma delimited). dan juga penyerdehanaan data pada dataset sehingga lebih mudah dalam pengolahannya.

5) Proses Elbow

Pada tahap ini adalah proses yang digunakan dalam analisis clustering untuk menentukan jumlah cluster yang optimal dalam suatu dataset. untuk rumus dari elbow sendiri yaitu :

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

Keterangan :

K = clster ke – c

Xi = jarak data obtyek ke – i

ck = pusat cluster ke – i

6) Proses Kmeans

Dalam langkah ini, model akan dibuat menggunakan teknik clustering untuk mengidentifikasi pola dalam proses data mining. teknik clustering yang akan digunakan adalah kmeans. sebelum membuat model, langkah pertama adalah menentukan jumlah cluster yang sesuai. Untuk menentukan jumlah cluster yang optimal, peneliti akan menggunakan metode elbow. proses pengolahan dilakukan menggunakan platform Jupyter Notebook

7) Evaluasi

Pada tahapan evaluasi terinspirasi dari hasil proses pengolahan data mining yang nantinya dihasilkan dan di uji pada tahap proses pemodelan sebelumnya. setelah melakukan tahap pemodelan, tahapan alur evaluasi akan dijalankan yaitu dapat dilakukan sebuah percobaan dilakukan untuk menentukan dan mengevaluasi jumlah K yang paling optimal dalam penelitian in dengan menggunakan silhouette.

8) Silhouette

dalam Tahap ini adalah proses yang digunakan dalam mengukur kelompok terbaik dari proses elbow apakah benar atau salah dalam menentukan kelompok terbaiknya. Untuk rumus dari silhouette adalah :

$$S = \frac{b-a}{\max(a,b)} \quad (2)$$

Keterangan

S = silhouette

a = jarak rata-rata sebuah sampel dengan semua titik dalam kelas yang sama

b = jarak rata-rata sebuah sampel dan semua titik dalam cluster terdekat

III. HASIL DAN PEMBAHASAN

A. Pembentukan Data

Proses pada halaman ini menjelaskan terkait proses alur yang nantinya akan dilakukan dalam hal pengerjaan yang telah di susun pada bab sebelumnya. dimana proses penyusunan bab ini menggunakan metode KDD (*Knowledge Discovery In Database*) sebagai alur dalam proses penelitian ini.

1) Data Creation

Pada tahap ini, peneliti memilih dan menyeleksi data yang akan digunakan pada saat data mining. Langkah ini sangat penting karena akan berpengaruh pada pengolahan data nantinya. pemilihan data yang tepat akan memastikan bahwa analisis yang dilakukan akurat dan relevan. dalam penelitian ini, atribut yang digunakan adalah kunjungan wisatawan nusantara dan mancanegara dari tahun 2018 sampai 2023. untuk contoh data dapat dilihat pada gambar 2.

NO	DAYA TARIK WISATA	JANUARI		PEBRUARI		MARET		APRIL	
		WISMAN	WISNU S	WISMAN	WISNU S	WISMAN	WISNU S	WISMAN	WISNU S
1	WANA WISATA PADUSAN PACE	-	54.170	-	32.103	-	35.066	-	46.792
2	AIR TERJUN DLUNDUNG	-	3.872	-	3.106	-	3.278	-	4.221
3	PETIRTAAN JOLOTUNDO	8	9.953	6	6.061	6	6.331	10	8.215
4	MAKAM TROLOYO	-	32.280	-	25.487	-	45.012	-	69.876
5	PEMANDIAN AIR PANAS	3	41.944	-	26.056	-	29.857	-	36.132
6	MUSEUM TROWULAN	182	9.258	156	9.153	133	6.836	127	4.229
7	EKOWISATA TANJUNGAN	-	24.121	-	12.659	-	10.122	-	10.888
8	SITI INGIL	-	0	-	-	-	62	-	211
9	COBAN CANGGU	-	4.350	-	2.343	-	2.328	-	3.416
10	PEMANDIAN UBALAN PACET	-	21.620	-	10.322	-	10.122	-	4.181

Gbr. 2 Data Kunjungan Wisata

2) Preprocessing atau Cleaning

Pada tahap ini kita akan memproses data yang sudah kita pilih atau kita seleksi pada tahap sebelumnya yaitu data selection. pada data mining kita membutuhkan data yang bersih dan tidak ada nilai null dalam data, dan untuk itu. diperlukan pembersihan data sehingga dapat digunakan pada proses data mining nantinya. untuk menghilangkan nilai null saya menggunakan sebuah Teknik untuk mengisi data yang bernilai null teknik tersebut dinamakan teknik imputasi. untuk hasil dari preprocessing hal ini dapat diamati pada tabel 1 :

TABEL I
HASIL PREPROCECING

Objek Wisata	18 jan	18 feb	18 mar	...
Wana Wisata Padusan	54170	32103	35066	...
Pemandian Air Panas	41944	26056	29857	...
Air Terjun Dlundung	3872	3106	3278	...
Petirtaan Jolotundo	9953	6061	6331	...
Makam Troloyo	32280	25487	45012	...
Museum Trowulan	9258	9153	6836	...
Ubalan Waterpark	21620	10322	10122	...
Coban Canggu	4350	2343	2328	...
Candi Brahu	3902	1824	2303	...
...	...	...	...	...

3) Data Integration

Pada tahap ini ditambahkan beberapa atribut yang nantinya akan mempermudah untuk melakukan analisis dan untuk atribut yang saya tambahkan yaitu jarak ke pusat kota, rating, penginapan dan kuliner. untuk hasil dapat dilihat pada tabel 2.

TABEL II
KUNJUNGAN WISATA

Objek Wisata	18 wisnu jan	18 wisnu feb	18 wisnu mar	...
Wana Wisata Padusan	54170	32103	35066	...
Pemandian Air Panas	41944	26056	29857	...
Air Terjun Dlundung	3872	3106	3278	...
Petirtaan Jolotundo	9953	6061	6331	...
Makam Troloyo	32280	25487	45012	...
Museum Trowulan	9258	9153	6836	...
Ubalan Waterpark	21620	10322	10122	...
Coban Cangu	4350	2343	2328	...
Candi Brahu	3902	1824	2303	...
Candi Bajang Ratu	4224	2174	2170	...
Candi tikus	4942	2655	2609	...
Taman Ganjaran	17886	11162	12007	...
...	...	...	...	...

Pada tabel 3 menunjukkan data dari tempat wisata ke pusat kota yang di peroleh dari hasil proses integrasi data melalui berbagai sumber yang valid.

TABEL III
JARAK KE PUSAT KOTA

Objek Wisata	jarak
Wana Wisata Padusan	32
Pemandian Air Panas	32
Air Terjun Dlundung	40
Petirtaan Jolotundo	34
Makam Troloyo	16
Museum Trowulan	14
Ubalan Waterpark	31
Coban Cangu	33
Candi Brahu	14
Candi Bajang Ratu	15
Candi tikus	15
Taman Ganjaran	38
...	...

Pada tabel 4 menunjukkan atau hasil dari proses proses olah yang di lakukan pada data integrasi utuk mendapatkan atribut rating yang akan digunakan dalam proses alur dari data mining. Dimana dari proses tersebut yang di dapat lewat berbagai sumber yang valid.

TABEL IV
RATING TEMPAT WISATA

Objek Wisata	Rating
Wana Wisata Padusan	4,4
Pemandian Air Panas	4,1
Air Terjun Dlundung	4,5
Petirtaan Jolotundo	4,7
Makam Troloyo	4,8
Museum Trowulan	4,5
Ubalan Waterpark	4,6
Coban Cangu	4,3
Candi Brahu	4,7
Candi Bajang Ratu	4,6
Candi tikus	4,5
Taman Ganjaran	4,6
...	...

Pada tabel 5 adalah proses hasil integrasi utuk mendapatkan atribut penginapan yang di dapat lewat berbagai seumber yang valid.

TABEL V
PENGINAPAN

Objek Wisata	Penginapan
Wana Wisata Padusan	1
Pemandian Air Panas	1
Air Terjun Dlundung	1
Petirtaan Jolotundo	1
Makam Troloyo	1
Museum Trowulan	1
Ubalan Waterpark	1
Coban Cangu	1
Candi Brahu	1
Candi Bajang Ratu	1
Candi tikus	1
Taman Ganjaran	1
...	...

Pada Tabel 6 ini adalah hasil proses integrasi untuk mendapatkan atribut kuliner yang digunakan dalam proses data mining yang di dapat lewat berbagai sumber yang valid

TABEL VI
KULINER

Objek Wisata	Kuliner
Wana Wisata Padusan	1
Pemandian Air Panas	1
Air Terjun Dlundung	1
Petirtaan Jolotundo	1
Makam Troloyo	1
Museum Trowulan	1
Ubalan Waterpark	1
Coban Cangu	1
Candi Brahu	1
Candi Bajang Ratu	1
...	...

4) Data Transformasi

Pada proses ini untuk data data pada atribut yang panjang akan di sederhanakan, hal itu dapat mempermudah dalam melakukan analisis yang lebih maksimal dan mengurangi kesalahan karena kepanjangan dataset yang digunakan. juga pengubahan file bentuk excel ke CSV sebagai syarat jalan nya pengkodean pada python. untuk hasil dari penyederhaan atribut dapat diamati pada tabel dibawah ini :

TABEL VII
HASIL DATA TRANSFORMASI

Objek Wisata	18 wisnu jan	...	rtng	Pgnp	klmr
Wana Wisata Padusan	54170	...	4,4	1	1
Pemandian Air Panas	41944	...	4,1	1	1
Air Terjun Dlundung	3872	...	4,5	1	1
Petirtaan Jolotundo	9953	...	4,7	1	1
Makam Troloyo	32280	...	4,8	1	1
Museum Trowulan	9258	...	4,5	1	1
Ubalan Waterpark	21620	...	4,6	1	1
Coban Cangu	4350	...	4,3	1	1
Candi Brahu	3902	...	4,7	1	1
...	...	...	...	...	...

B. Penerapan K-Means Dari Data Yang Terbentuk

Penelitian ini mengambil data dari tahun 2018 sampai 2023 dan juga dari website dimana atributnya berisi jumlah kunjungan wisata nusantara, kunjungan wisatawan mancanegara, jarak ke pusat kota, rating, penginapan, dan kuliner. proses pengolahan data mining menggunakan metode kmeans menggunakan tool python. untuk proses perhitungan dapat diamati pada tahapan berikut ini :

1) Proses Menginput Library

Sklearn.cluster dipakai untuk melakukan k-means clustering, sklearn.metrics dipakai untuk mengevaluasi skor siluet, numpy dipakai untuk memanipulasi dan mengolah array data, pandas dipakai untuk membaca dataset yang akan diproses, sklearn.preprocessing dipakai untuk standarisasi data, dan matplotlib.pyplot dipakai untuk menampilkan grafik saat melakukan proses elbow.

2.) Proses Pembacaan Dataset Berformat CSV Yang Akan Di olah

```
df= pd.read_csv("datashetfix.csv")
df.head(10)
```

ek Wisata	18 wisnu jan	18 wisnu feb	18 wisnu mar	18 wisnu ap	18 wisnu mei	18 wisnu jun	18 wisnu jul	18 wisnu ag	18 wisnu sep
A WISATA PADUSAN	54170	32103	35066	46792	32995	54302	54028	37116	55163
KOLAM ADIAN AIR PANAS	41944	26056	29857	36132	22663	43931	40244	27951	39090
AIR TERJUN DLUNDUNG	3872	3106	3278	4221	3280	5698	5809	5895	4321
PETIRTAAN JOTUNDO	9953	6061	6331	8215	5304	6903	8429	7282	9080
MAKAM TROLOYO	32280	25487	45012	69876	47236	14489	33550	19746	41360
MUSEUM TROWULAN	9258	9153	6836	4229	2240	2162	3049	1954	2286
UBALAN WATERPARK	21620	10322	10122	4181	3269	11750	6189	2918	3369
COBAN CANGGU	4350	2343	2328	3416	2215	4357	3537	1789	3202
CANDI BRAHU	3902	1824	2303	1826	975	1163	1342	956	1057
CANDI BAJANG RATU	4224	2174	2170	2032	1007	1102	1253	885	954

Gbr. 3 Hasil Pembacaan Dataset

3) Proses Standarisasi Data Pada Data

```
data_sceale = preprocessing.scale(data)
data_sceale
```

```
array([[ 4.0152697,  3.76904038,  3.17125938, ...,  0.01375684,
        0.81649658,  0.57735027],
       [ 3.0093975,  2.97641848,  2.64118294, ..., -0.39894828,
        0.81649658,  0.57735027],
       [-0.12290794, -0.03179597, -0.06353997, ...,  0.15132521,
        0.81649658,  0.57735027],
       ...,
       [-0.44146979, -0.43892077, -0.39711467, ..., -0.12381153,
        0.81649658,  0.57735027],
       [ 1.54304218,  1.22038148,  0.63291676, ..., -0.26137991,
        0.81649658,  0.57735027],
       [-0.44146979, -0.43892077, -0.39080544, ...,  0.28889358,
        0.81649658,  0.57735027]])
```

Gbr. 3 Hasil Standarisasi Data

4) Proses pengolahan K-means Menggunakan Metode Elbow Dengan Inisiasi K-Means++ dan Inisiasi Random

Metode Elbow

```
Inertia = []
for n_clusters in range_n_clusters:
    kmeans = KMeans(n_clusters=n_clusters,
                    random_state=0) kmeans.fit(X)
    inertia.append(kmeans.inertia_)
```

Inisialisasi Kmeans++

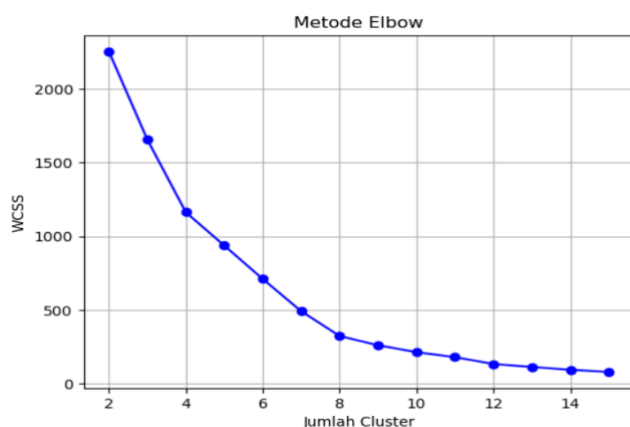
```
inertia = []
for i in range(2, 16):
    kmeans = KMeans(n_clusters=i, init='k-means++',
                    max_iter=300, n_init=16, random_state=123)
    kmeans.fit(data_sceale)
    wcss.append(kmeans.inertia_)
```

Inisialisasi Random

```
inertia = []
for k in range(2, 16):
    kmeans = KMeans(n_clusters=k, init='random')
    kmeans.fit(data_sceale)
    inertia.append(kmeans.inertia_)
```

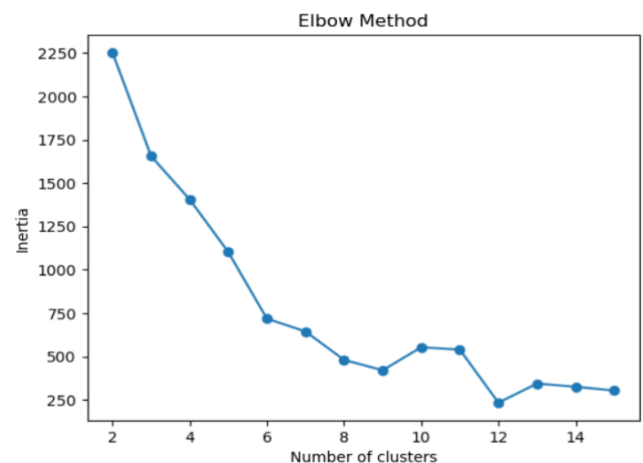
5) Proses Menampilkan Hasil K-means Inisiasi K-Means++ Dan Inisiasi Random Menggunakan Elbow

Hasil Elbow kmeans++



Gbr. 4 elbow kmeans++

Hasil Elbow inisialisasi Random



Gbr. 5 Elbow Random

Menurut hasil dari clustering menggunakan elbow maka dapat dilihat untuk K terbaik adalah 2 cluster. dimana dapat dilihat dari grafik dari 3 maka jalur arah grafik semakin berdempetan sejajar. untuk memastikan apakah k terbaik adalah 2 bisa dilihat pada tahap selanjutnya yaitu evaluasi apakah cluster 2 adalah terbaik apa tidak.

C. Skenario Pengujian K-Means Dan Hasil Pembahasan

Pada proses tahap evaluasi hasil dari perhitungan kmeans akan di evaluasi menggunakan silhouette. metode analisis ini dapat menentukan jumlah cluster yang paling optimal dan untuk percobaan akan menggunakan inisialisasi Kmeans++ dan inisialisasi random. data yang gunakan masih sama mengambil dari data kunjungan wisatawan nusantara, mancanegara, jarak ke pusat kota, rating, penginapan dan kuliner. jumlah cluster yang akan di uji coba yaitu k = 2 sampai k= 15.

1) Pengujian K-Means Dengan Inisialisasi KMeans++

Untuk pengujian Kmeans++ ini akan dilakukan sebanyak 14 kali dari 14 kali percobaan K terbaik berapa yang akan di pakai dalam pengelompokan untuk hasil dari penelitian skripsi ini. Untuk hasil dari pengujian dapat diamati dari tabel 8.

TABEL VIII
HASIL SILHOUTE INISIALISASI KMEANS++

K terbaik	Hasil silhoute
2	0.6817446298183374
3	0.674182646976267
4	0.6711545464660984
5	0.40826009296940596
6	0.3799089493718147
7	0.4346887503877726
8	0.4346887503877726
9	0.4346887503877726
...	...

Dari pengujian di atas maka di dapat K terbaik adalah K=2 oleh karena itu maka akan di lakukan pengelompokan menjadi 2 kelompok.

2) Pengujian K-Means Dengan Inisialisasi Random

Untuk pengujian Kmeans dengan inisialisasi random ini sama seperti inisialisasi Kmeans++ akan di lakukan sebanyak 14 dan mencari K terbaiknya dari 2 sampai 15. untuk hasil dari pengujian dapat dimati di tabel 9.

TABEL IX
HASIL SILHOUTE INISIALISASI RANDOM

K terbaik	Hasil silhouette
2	0.6817446298183374
3	0.674182646976267
4	0.45616571367656145
5	0.47666087848574473
6	0.4410591882198247
7	0.24094647038318345
8	0.4235440791536046
9	0.4159497205900018
...	...

Dari kedua pengujian di atas maka di dapat k terbaik adalah k=2 oleh karena itu maka akan di lakukan pengelompokan menjadi 2 kelompok dan dapat dilihat pada tabel 10.

TABEL X
HASIL PENGUJIAN KMEANS ++ DAN RANDOM

No	Objek wisata	18 wisnu jan	...	kuliner	cluster
1	Wana Wisata Padusan	54170	...	1	1
2	Pemandian Air Panas	41944	...	1	1
3	Air Terjun Dlundung	3872	...	1	0
4	Petirtaan Jolotundo	9953	...	1	0
5	Makam Troloyo	32280	...	1	1
6	Museum Trowulan	9258	...	1	0
7	Ubalan Waterpark	21620	...	1	0
8	Coban Cangu	4350	...	1	0
9	Candi Brahu	3902	...	1	0
10	Candi Bajang Ratu	4224	...	1	0

Dari kedua pengujian di atas baik dari kmeans++ ataupun inisialisasi random memiliki hasil silhouette yang paling tinggi adalah k = 2 untuk percobaanya sendiri yang dipergunakan adalah k = 2 sampai k = 15. hasil dari k = 1 sebanyak 37 Wisata yang memiliki rata rata pengunjungnya terbilang standart ada juga yang pengunjungnya sedikit, untuk kuliner, penginapan di beberapa tempat wisata tidak ada. wisatanya antara lain adalah air terjun dlundung, petirtaan jolotundo, museum trowulan, ubalan waterpark, coban cangu, candi

brahu, candi bajang ratu, candi tikus, taman ganjaran, sumer gempong, out bound obech, bernah de valei, klurak eco park, ww suradadu, putuk shiwur, brambang outdoor selfie, indereng genitri, claket adventure, outbound obis, lumbung wit witan, perkemahan kali maron dewasa, perkemahan kali maron anak, pendakian tamiajeng, pendakian seloliman, bukit krapyak, welirang penanggungan adventure, alas Pelangi, kico bike park, air terjun grenjengan, pendakian puthuk gragal, cemara surya alam, pendakian lorokan, alas veenuz trawas, bukit kayoe putih, Parimas waterpark, ekowisata tanjungan, siti inggil. sedangkan k = 2 sebanyak 3 wisata memiliki rata rata pengunjung tinggi untuk kuliner, penginapan, ada di sekitar tempat wisata tersebut antarlain adalah kolam pemandian air panas, wana wisata padusan dan makam troloyo.

IV. KESIMPULAN

Berdasarkan dua kali uji coba yang menggunakan inisialisasi random dan inisialisasi kmeans++ dengan total 14 kali percobaan, diperoleh hasil silhouette tertinggi sebesar 0.6817446298183374. Data atribut yang digunakan meliputi kunjungan wisatawan mancanegara, wisatawan nusantara, rating, penginapan, kuliner, dan jarak ke pusat kota. Dari analisis ini, diperoleh nilai k terbaik sama dengan 2. Hasil pengelompokan menggunakan K-Means clustering menunjukkan bahwa klaster 1 terdiri dari 37 objek wisata, sedangkan klaster 2 terdiri dari 3 objek wisata. Contoh wisata di klaster 1 adalah parimas waterpark, candi tikus, dan candi brahu. Sedangkan, klaster 2 yang termasuk didalamnya adalah pemandian air panas, wana wisata padusan dan makam troloyo..

V. SARAN

Untuk penelitian selanjutnya, disarankan agar data yang digunakan diperbanyak karena banyak objek wisata di Mojokerto yang belum terdata dalam penelitian ini. selanjutnya untuk menerapkan fitur seleksi dalam analisis data pariwisata mojokerto. hal ini tidak hanya akan meningkatkan kualitas informasi yang diterima oleh para wisatawan, tetapi juga dapat meningkatkan efisiensi dan keakuratan analisis data..

UCAPAN TERIMA KASIH

dengan pengakuan dan rasa berterimakasih kepada Allah SWT atas segala petunjuk-Nya sehingga saat ini artikel saya yang berjudul "Implementasi Data Mining Untuk Pengelompokan Kunjungan Wisata Di Kabupaten Mojokerto" dapat diselesaikan tepat waktu dengan baik. penulis dengan rendah hati ingin menyampaikan ucapan terima kasih kepada orang tua yang sangat saya cintai, yang gigih dalam memberikan doa dan memberi dorongan ataupun petunjuk kepada peneliti. Ibu Dr. Yuni Yamasari, S.Kom., M.Kom. Sebagai dosen pembimbing skripsi saya yang telah memberikan ilmu dan bimbingannya. bapak Aditya Prapanca, S.T., M.Kom. dan Bapak I Made Suartana, S.Kom., M.Kom. sebagai dosen yang telah memberi banyak sekali masukan sehingga skripsi saya menjadi lebih sempurna. kepada sahabat

skripsi yang telah memberi banyak dukungan dalam perangkin penelitian ini.

REFERENSI

- [1] M. Tobing, "Pengaruh Jumlah Obyek Wisata, Tingkat Penghunian Kamar, Dan Jumlah Kunjungan Wisatawan Terhadap Pendapatan Asli Daerah Kabupaten Simalungun," *Jurnal Ekuilnomi*, vol. 3, no. 2, pp. 127–139, Dec. 2021, doi: 10.36985/ekuilnomi.v3i2.263.
- [2] A. Dwijulianto and P. A. R. Dewi, "Manajemen Komunikasi Pariwisata 'The Spirit of Majapahit' Di Kabupaten Mojokerto," *Commercium*, vol. 2, no. 1, pp. 29–32, 2019, [Online]. Available: <https://jurnalmahasiswa.unesa.ac.id/index.php/Commercium/article/view/29173/26716>
- [3] B. M. Al-Fahmi, E. Rahmawati, and T. Sagirani, "Penerapan K-Means Clustering Pada Pariwisata Kabupaten Bojonegoro Untuk Mendukung Keputusan Strategi Pemasaran," *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 9, no. 2, pp. 141–149, Aug. 2023, doi: 10.25077/teknosi.v9i2.2023.141-149.
- [4] L. Maulida, "Penerapan Datamining Dalam Mengelompokkan Kunjungan Wisatawan Ke Objek Wisata Unggulan Di PROV. DKI Jakarta Dengan K-Means," *JISKA (Jurnal Informatika Sunan Kalijaga)*, vol. 2, no. 3, p. 167, 2018, doi: 10.14421/jiska.2018.23-06.
- [5] H. Hasanah, N. A. Sudibyo, and R. M. Galih, "Data Mining Using K-Means Clustering Algorithm for Grouping Countries of Origin of Foreign Tourist," in *Nusantara Science and Technology Proceedings*, Galaxy Science, Sep. 2021. doi: 10.11594/nstp.2021.1112.
- [6] M. Egypt Pratama and A. Finandhita, "Penerapan Metode Clustering Untuk Pengelompokan Potensi Wisata Di Kabupaten Sumedang," *Skripsi. Universitas KOMPuter Indonesia*, pp. 1–8, 2019.
- [7] Sari and D. Hartama, *Data Mining: Algoritma K-Means Pada Pengelompokan Wisata Asing ke Indonesia Menurut Provinsi*. 2018. [Online]. Available: <http://seminar-id.com/semnas-sensasi2018.htmlPage%7C322>
- [8] F. N. Dhewayani *et al.*, "Implementasi K-Means Clustering untuk Pengelompokan Daerah Rawan Bencana Kebakaran Menggunakan Model CRISP-DM," *Jurnal Teknologi dan Informasi*, 2022, doi: 10.34010/jati.v12i1.
- [9] Han, Jiawei, Kamber, and Micheline, "Data Mining: Concepts and Techniques Second Edition," 2022.
- [10] S. Naning, R. Pranata, A. N. M. Clara, and O. S. Rahajeng, "Pengelompokan Kunjungan Wisata Kabupaten Kulon Progo Tahun 2019 Menggunakan K-Means Clustering," *JUST IT: Jurnal Sistem ...*, vol. 12, no. 1, pp. 38–45, 2022, [Online]. Available: <https://jurnal.umj.ac.id/index.php/just-it/article/view/8249%0Ahttps://jurnal.umj.ac.id/index.php/just-it/article/download/8249/7242>
- [11] I. Wahyudi, M. B. Sulthan, and L. Suhartini, "Analisa Penentuan Cluster Terbaik Pada Metode K-Means Menggunakan Elbow Terhadap Sentra Industri Produksi Di Pamekasan," *Jurnal Aplikasi Teknologi Informasi dan Manajemen (JATIM)*, vol. 2, no. 2, pp. 72–81, 2021, doi: 10.31102/jatim.v2i2.1274.
- [12] H. Alexander, Y. Umaidah, and M. Jajuli, "Implementasi Clustering Untuk Menentukan Efektifitas Nilai Siswa Sesudah Pandemi Covid-19 Menggunakan Algoritma K-Means," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 7, no. 3, pp. 1493–1500, 2023, doi: 10.36040/jati.v7i3.7174.
- [13] S. Paembonan and H. Abduh, "Penerapan Metode Silhouette Coefficient untuk Evaluasi Clustering Obat," *PENA TEKNIK: Jurnal Ilmiah Ilmu-Ilmu Teknik*, vol. 6, no. 2, p. 48, 2021, doi: 10.51557/pt_jiit.v6i2.659.
- [14] A. Winarta and W. J. Kurniawan, "Optimasi Cluster K-Means Menggunakan Metode Elbow Pada Data Pengguna Narkoba Dengan Pemrograman Python," *JTIK (Jurnal Teknik Informatika Kaputama)*, vol. 5, no. 1, pp. 113–119, 2021, doi: 10.59697/jtik.v5i1.593.