



**IMPLEMENTASI DATA MINING MENGGUNAKAN *K-MEANS*
CLUSTERING DALAM Mendukung PENGELOMPOKAN TEMPAT
WISATA KABUPATEN BOJONEGORO**



UNIVERSITAS
Dinamika

Oleh:
BARRON MAHARDHIKA AL-FAHMI
19410100107

FAKULTAS TEKNOLOGI DAN INFORMATIKA
UNIVERSITAS DINAMIKA
2023

**IMPLEMENTASI DATA MINING MENGGUNAKAN *K-MEANS*
CLUSTERING DALAM Mendukung PENGELOMPOKAN TEMPAT
WISATA KABUPATEN BOJONEGORO**

TUGAS AKHIR

**Diajukan sebagai salah syarat untuk menyelesaikan
Program Sarjana**



**UNIVERSITAS
Dinamika**

Oleh:

**Nama : Barron Mahardhika Al-Fahmi
NIM : 19410100107
Program Studi : S1 Sistem Informasi**

**FAKULTAS TEKNOLOGI DAN INFORMATIKA
UNIVERSITAS DINAMIKA**

2023

Tugas Akhir

IMPLEMENTASI DATA MINING MENGGUNAKAN *K-MEANS* *CLUSTERING* DALAM Mendukung Pengelompokan Tempat Wisata Kabupaten Bojonegoro

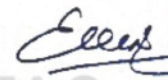
Dipersiapkan dan disusun oleh
Barron Mahardhika Al-Fahmi
NIM: 19410100107

Telah diperiksa, dibahas dan disetujui oleh Dewan Pembahas
Pada: 13 Juli 2023

Susunan Dewan Pembahas

Pembimbing

- I. Endra Rahmawati, M.Kom.
NIDN. 0712108701
- II. Tri Sagirani, S.Kom., M.MT.
NIDN. 0731017601



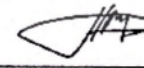
Digitally signed by
Endra Rahmawati
Date: 2023.08.04
08:35:34 +07'00'



Digitally signed by
Universitas Dinamika
Date: 2023.08.04
10:51:59 +07'00'

Pembahas

- I. Tutut Wuriyanto, M.Kom.
NIDN. 0703056702



Date: 2023.08.07
10:47:28 +07'00'

Tugas Akhir ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana:



Digitally signed by
Universitas Dinamika
Date: 2023.08.08
08:45:26 +07'00'

Tri Sagirani, S.Kom., M.MT.
NIDN. 0731017601

Dekan Fakultas Teknologi dan Informatika
UNIVERSITAS DINAMIKA

Tidak ada yang terlalu cepat ataupun terlambat. Semua berjalan sesuai dengan ketentuan dan takdir yang tepat. Selalu yakin dan bersyukur bahwa rencana Allah sangatlah luar biasa.

- Barron Mahardhika Al-Fahmi -



UNIVERSITAS
Dinamika



*Kupersembahkan kepada
Keluarga,
Bapak Ibu Dosen,
Civitas Universitas Dinamika
Teman, sahabat dan almamater
Universitas Dinamika yang kubanggakan.*

SURAT PERNYATAAN
PERSETUJUAN PUBLIKASI DAN KEASLIAN KARYA ILMIAH

Sebagai mahasiswa Universitas Dinamika, Saya :

Nama : Barron Mahardhika Al-Fahmi
NIM : 19410100107
Program Studi : S1 Sistem Informasi
Fakultas : Fakultas Teknologi dan Informatika
Jenis Karya : Tugas Akhir
Judul Karya : **IMPLEMENTASI DATA MINING MENGGUNAKAN
K-MEANS CLUSTERING DALAM Mendukung
PENGELOMPOKAN TEMPAT WISATA
KABUPATEN BOJONEGORO**

Menyatakan dengan sesungguhnya bahwa :

1. Demi pengembangan Ilmu Pengetahuan, Teknologi dan Seni, Saya menyetujui memberikan kepada Universitas Dinamika Hak Bebas Royalti Non-Eksklusif (*Non-Exclusive Royalty Free Right*) atas seluruh isi/sebagian karya ilmiah Saya tersebut diatas untuk disimpan, dialihmediakan, dan dikelola dalam bentuk pangkalan data (*database*) untuk selanjutnya didistribusikan atau dipublikasikan demi kepentingan akademis dengan tetap mencantumkan nama Saya sebagai penulis atau pencipta dan sebagai pemilik Hak Cipta.
2. Karya tersebut diatas adalah hasil karya asli Saya, bukan plagiat baik sebagian maupun keseluruhan. Kutipan, karya, atau pendapat orang lain yang ada dalam karya ilmiah ini semata-mata hanya sebagai rujukan yang dicantumkan dalam Daftar Pustaka Saya.
3. Apabila dikemudian hari ditemukan dan terbukti terdapat tindakan plagiasi pada karya ilmiah ini, maka Saya bersedia untuk menerima pencabutan terhadap gelar kesarjanaan yang telah diberikan kepada Saya.

Demikian surat pernyataan ini saya buat dengan sebenarnya.

Surabaya, 27 Juni 2023



Barron Mahardhika Al-Fahmi
NIM : 19410100107

ABSTRAK

Dinas Kebudayaan dan Pariwisata Kabupaten Bojonegoro adalah suatu lembaga pemerintahan yang bergerak dalam bidang pemberdayaan pariwisata yang ada di Kabupaten Bojonegoro. Pada dinas ini terdapat beberapa bidang dalam menjalankan pengelolaan pariwisata yaitu Bidang Pariwisata, Bidang Kebudayaan, dan Bidang Ekonomi Kreatif. Pada bidang pariwisata memiliki tujuan untuk meningkatkan kontribusi sektor pariwisata di Kabupaten Bojonegoro. Dari 25 tempat wisata yang terdapat di Kabupaten Bojonegoro belum dilakukan sebuah pengelompokan tempat wisata. Dalam upaya memajukan wisata salah satu faktor yang berpengaruh adalah kunjungan wisatawan pada tempat-tempat pariwisata. Masing-masing tempat pariwisata mempunyai karakteristik yang berbeda-beda, sehingga hal ini akan menimbulkan preferensi tersendiri dalam berwisata. Salah satu metode yang dapat digunakan untuk melakukan pengelompokan tempat wisata adalah *k-means clustering*. Diantara sekian metode *clustering* yang ada, metode *k-means clustering* merupakan yang paling sederhana dan umum digunakan. Konsep dari algoritma ini yaitu obyek yang merupakan anggota dari suatu kumpulan data, dikelompokkan menjadi sejumlah *k cluster* atau kelompok. Dalam menentukan banyaknya nilai *k* dapat menggunakan sebuah metode yang disebut *elbow*. Metode *elbow* digunakan untuk menentukan jumlah kelompok yang optimal dalam *k-means clustering*. Pada penelitian ini telah dilakukan uji coba dari 10 data, 15 data, dan 25 data. Dari hasil uji coba metode *elbow* dengan menghitung nilai *Sum of Square Error* mendapatkan hasil *K* yang optimal berada pada $K=3$ yang berarti pengelompokan tempat wisata dapat dibagi menjadi 3 kelompok. Setelah diidentifikasi karakteristik anggota dari setiap *cluster*, selanjutnya adalah proses pembuatan visualisasi *dashboard* agar lebih informatif dan mudah dibaca oleh pengguna. Hal ini akan membantu Dinas dalam mengembangkan pengalaman wisata, menyediakan layanan yang lebih sesuai dengan preferensi pengunjung, dan pengembangan destinasi wisata.

Kata Kunci: *K-Means Clustering*, Metode *Elbow*, *SSE (Sum of Square Error)*, *Dashboard*

KATA PENGANTAR

Puji syukur ke hadirat Allah SWT yang telah memberikan rahmat dan hidayah-Nya sehingga penulis mampu menyelesaikan Tugas Akhir dengan judul “Implementasi Data Mining Menggunakan *K-Means Clustering* Dalam Mendukung Pengelompokan Tempat Wisata Kabupaten Bojonegoro”. Laporan Tugas Akhir ini disusun sebagai salah satu syarat menyelesaikan program strata satu di Universitas Dinamika.

Penyelesaian Tugas Akhir ini tidak terlepas dari bantuan berbagai pihak yang telah memberikan banyak masukan, nasihat, saran, kritik, dan dukungan moral maupun materil kepada penulis. Oleh karena itu penulis menyampaikan rasa terima kasih kepada:

1. Mama, Papa dan keluarga tercinta keluarga yang selalu mendoakan, mendukung, dan memberikan semangat di setiap langkah dan aktivitas penulis.
2. Bapak Dr. Anjik Sukmaaji, S.Kom., M.Eng. selaku Ketua Program Studi S1 Sistem Informasi Universitas Dinamika yang telah memberikan izin kepada penulis untuk melaksanakan Tugas Akhir.
3. Ibu Endra Rahmawati, M.Kom. selaku Dosen S1 Sistem Informasi sekaligus dosen pembimbing pertama dalam kegiatan Tugas Akhir yang senantiasa meluangkan waktu dan pikirannya untuk membimbing, mendukung, dan memberikan arahan kepada penulis dalam proses Tugas Akhir.
4. Ibu Tri Sagirani, S.Kom., M.MT. selaku Dosen S1 Sistem Informasi sekaligus dosen pembimbing kedua dalam kegiatan Tugas Akhir yang senantiasa sabar dan tekun dalam mengarahkan penulis.
5. Bapak Tutut Wuriyanto, M.Kom. selaku Dosen Penguji yang telah memberikan saran serta masukan dalam penyelesaian Tugas Akhir ini.
6. Dinas Kebudayaan dan Pariwisata Kabupaten Bojonegoro yang merupakan instansi pendukung untuk menyelesaikan tugas akhir ini.
7. Teman-teman tercinta yang memberikan bantuan dan dukungan dalam menyelesaikan Tugas Akhir ini.
8. Pihak-pihak lain yang tidak dapat disebutkan satu-persatu yang telah memberikan bantuan dan dukungan kepada penulis.

9. *Last but not least, I wanna thank me, for believing in me, for doing all this hard work, for having no days off, for never quitting, for just being me at all times.*

Semoga Allah SWT memberikan balasan yang setimpal kepada semua pihak yang telah membantu dan memberikan bimbingan serta nasehat dalam proses menyelesaikan Tugas Akhir ini.

Penulis menyadari bahwa Tugas Akhir yang dikerjakan ini masih banyak terdapat kekurangan sehingga kritik yang bersifat membangun dan saran dari semua pihak sangatlah diharapkan agar dapat diperbaiki menjadi lebih baik lagi. Semoga Tugas Akhir ini dapat diterima dan bermanfaat bagi penulis dan semua pihak.

Surabaya, 30 Juni 2023



UNIVERSITAS
Dinamika

Penulis

DAFTAR ISI

	Halaman
ABSTRAK	vii
KATA PENGANTAR.....	viii
DAFTAR ISI.....	x
DAFTAR TABEL	xii
DAFTAR GAMBAR.....	xiii
DAFTAR LAMPIRAN.....	xiv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah.....	3
1.4 Tujuan	4
1.5 Manfaat	4
BAB II LANDASAN TEORI	5
2.1 Penelitian Terdahulu	5
2.2 Data Mining	7
2.3 <i>K-Means</i>	8
2.4 <i>Phyton</i>	10
2.5 Metode <i>Elbow</i>	11
2.6 <i>Power BI</i>	12
BAB III METODOLOGI PENELITIAN	14
3.1 Tahap Awal	14
3.1.1 Studi Literatur	14
3.1.2 Pengumpulan Data	15
3.2 Tahap Pengembangan	15
3.2.1 <i>Business Understanding</i>	15
3.2.2 <i>Data Understanding</i>	15
3.2.3 <i>Data Preparation</i>	16
3.2.4 <i>Modelling</i>	19
3.3 Tahap Akhir	22
3.3.1 <i>Evaluation</i>	22

BAB IV HASIL DAN PEMBAHASAN	24
4.1 Data Preparation.....	24
4.1.1 Hasil Pengolahan Data Kunjungan Wisata	24
4.1.2 Hasil Pengolahan Data Luas Wilayah Pariwisata.....	26
4.2 Modelling	27
4.3 <i>Evaluation</i>	31
4.3.1 Uji Coba 1 dengan 10 Data Tempat Wisata.....	31
4.3.2 Uji Coba 2 dengan 15 Data Tempat Wisata.....	32
4.3.3 Uji Coba 3 dengan 25 Data Tempat Wisata.....	33
4.3.4 Perhitungan <i>K-Means Clustering</i> $K=3$	34
4.3.5 Pembuatan Visualisasi Data dengan <i>Power BI</i>	38
BAB V PENUTUP.....	40
5.1 Kesimpulan	40
5.2 Saran.....	40
DAFTAR PUSTAKA.....	41
LAMPIRAN.....	43



DAFTAR TABEL

Halaman

Tabel 2.1 Penelitian Terdahulu	5
Tabel 3.1 Data yang Dibutuhkan	16
Tabel 3.2 Data Jumlah Kunjungan Wisata.....	17
Tabel 3.3 Data Kapasitas Kunjungan Tempat Wisata	18
Tabel 3.4 Data Luas Wilayah Tempat Wisata	18
Tabel 3. 6 Contoh Dataset.....	19
Tabel 3.7 Titik <i>Centroid</i> Awal.....	20
Tabel 3.8 Contoh Hasil Perhitungan Tiap <i>Cluster</i>	21
Tabel 3.9 Contoh Titik <i>Cluster</i> Baru	22
Tabel 4.1 Data Rata-Rata Kunjungan	25
Tabel 4.2 Persentase Jumlah Kunjungan	26
Tabel 4.3 Data Luas Wilayah Pariwisata	26
Tabel 4.4 Hasil <i>Insert Dataset</i>	28
Tabel 4.6 Hasil Pelabelan Setelah Proses <i>Clustering</i>	29
Tabel 4.7 Jumlah Anggota Dari Setiap <i>Cluster</i>	30
Tabel 4.8 Hasil Perhitungan <i>Sum of Square Error</i> 10 Data.....	31
Tabel 4.9 Hasil Perhitungan <i>Sum of Square Error</i> 15 Data.....	32
Tabel 4.10 Hasil Perhitungan <i>Sum of Square Error</i> 25 Data.....	33
Tabel 4.11 Hasil <i>Insert Dataset</i>	35
Tabel 4.13 Hasil Pelabelan Setelah Proses <i>Clustering</i>	35
Tabel 4.14 Jumlah Anggota Dari Setiap <i>Cluster</i>	37
Tabel 4.15 Hasil Rekap Anggota <i>Cluster</i>	37
Tabel 4.16 Karakteristik Anggota Setiap <i>Cluster</i>	37
Tabel L.1.1 Hasil Pelabelan Data K=2 sampai K=10	43
Tabel L.2.1 Hasil Pelabelan Data K=3	44
Tabel L.3.1 Data Jumlah Kunjungan Wisata	45
Tabel L.3.2 Data Rata-Rata Kunjungan.....	47

DAFTAR GAMBAR

	Halaman
Gambar 2.1 Metode CRISP-DM.....	7
Gambar 2.2 <i>Flowchart</i> Proses <i>K-Means Clustering</i>	10
Gambar 3.1 Metodologi Penelitian	14
Gambar 3.2 Contoh Diagram Dashboard.....	23
Gambar 4.1 Hasil <i>Scatter</i>	30
Gambar 4.2 Grafik <i>Sum of Square Error</i> 10 Data	32
Gambar 4.3 Grafik <i>Sum of Square Error</i> 15 Data	33
Gambar 4.4 Grafik <i>Sum of Square Error</i> 25 Data	34
Gambar 4.5 Hasil <i>Scatter</i>	36
Gambar 4.6 <i>Dashboard</i> Hasil Pengolahan Data dengan $K=3$	39



UNIVERSITAS
Dinamika

DAFTAR LAMPIRAN

	Halaman
Lampiran 1. Hasil Olah Data Pada Tahapan Modelling dengan Nilai $K=2$ sampai $K=10$	43
Lampiran 2. Hasil Olah Data Pada Tahapan Evaluation dengan Nilai $K=3$	44
Lampiran 3. Data yang Digunakan Pada Tahapan <i>Data Preparation</i>	45
Lampiran 4. Hasil Plagiarisme	49
Lampiran 5. Kartu Bimbingan	56
Lampiran 6. Biodata Penulis	58



UNIVERSITAS
Dinamika

BAB I

PENDAHULUAN

1.1 Latar Belakang

Dinas Kebudayaan dan Pariwisata Kabupaten Bojonegoro adalah suatu lembaga pemerintahan yang bergerak dalam bidang pemberdayaan pariwisata yang ada di Kabupaten Bojonegoro. Banyak sekali pariwisata yang ada di Bojonegoro seperti Dander Waterpark, Kayangan Api, Waduk Pacal, Negeri Atas Angin, Agrowisata Belimbing, Agrowisata Jambu Padang, Wisata Edukasi Gerabah, Teksas Wonocolo, Growgoland, dan masih banyak lagi. Pada dinas ini terdapat beberapa bidang dalam menjalankan pengelolaan pariwisata yaitu Bidang Pariwisata, Bidang Kebudayaan, dan Bidang Ekonomi Kreatif. Pada bidang pariwisata memiliki tujuan untuk meningkatkan kontribusi sektor pariwisata di Kabupaten Bojonegoro. Perkembangan pariwisata menjadi salah satu dari bagian pembangunan ekonomi agar dapat meningkatkan pertumbuhan ekonomi dalam suatu daerah. Perkembangan pariwisata sendiri di suatu daerah memberikan dampak positif terhadap ekonomi dan menjadikannya salah satu sumber pendapatan daerah (Hanafi Ahmad, 2022). Pendapatan daerah dapat diperoleh melalui pembangunan dan pendayagunaan sumber daya serta pariwisata.

Pembangunan pariwisata sangatlah penting, karena pariwisata merupakan potensi yang menjadi pendorong kehadiran wisatawan ke suatu daerah tujuan. Hal tersebut dapat direalisasikan dengan cara melakukan perluasan, pembangunan potensi wisata dan pemerataan daya tarik wisata. Guna mendukung terealisasinya hal tersebut perlu dibuat sebuah pengelompokan tempat wisata yang dilihat dari sisi pemanfaatan luas potensi wisata dan jumlah kunjungan dengan tujuan mengetahui tempat wisata mana saja yang tergolong sebagai *high priority* (prioritas tinggi dalam hal pengembangan tempat wisata dengan jumlah kunjungan yang tinggi), *high potential* (tempat wisata yang memiliki potensi tinggi dalam pengelolaan tempat wisata dan menarik kunjungan wisatawan), dan *low potential* (tempat wisata yang masih memiliki potensi rendah dalam hal menarik kunjungan wisatawan). Pengelompokan ini juga bisa menjadi evaluasi diri bagi pengelola

tempat wisata dengan memahami posisi *cluster* sehingga pengelola akan berusaha melalui strategi-strateginya untuk dapat naik level/cluster. Dari 25 tempat wisata yang terdapat di Kabupaten Bojonegoro belum dilakukan sebuah pengelompokan tempat wisata. Dalam upaya memajukan wisata salah satu faktor yang berpengaruh adalah kunjungan wisatawan pada tempat-tempat pariwisata. Masing-masing tempat pariwisata mempunyai karakteristik yang berbeda-beda, sehingga hal ini akan menimbulkan preferensi tersendiri dalam berwisata.

Pengelompokan tempat wisata merupakan salah satu hal yang penting dalam industri pariwisata. Dengan memahami karakteristik dan preferensinya para penyedia layanan pariwisata dapat mengembangkan destinasi wisata (Imam Iskandar & Ariffudin Islam, 2021). Salah satu metode yang dapat digunakan untuk melakukan pengelompokan tempat wisata adalah *k-means clustering*. Diantara sekian metode *clustering* yang ada, metode *k-means clustering* merupakan yang paling sederhana dan umum digunakan. Konsep dari algoritma ini yaitu obyek yang merupakan anggota dari suatu kumpulan data, dikelompokkan menjadi sejumlah *k cluster* atau kelompok. Sehingga dalam aplikasinya, nilai *k* atau jumlah *cluster*nya harus ditentukan terlebih dahulu (Harahap, 2017). Secara detail, dalam penerapannya, ukuran ketidakmiripan biasa digunakan dalam rangka mengelompokkan obyek-obyek ini ke dalam suatu *cluster*. Jika dua obyek atau data titik mempunyai jarak cukup dekat, maka dua obyek itu mirip (H. Witten et al., 2011). Hal ini mengindikasikan bahwa, semakin dekat jarak antar dua obyek, maka semakin tinggi tingkat kemiripannya. Semakin besar nilai jarak, semakin tinggi tingkat ketidakmiripannya.

Dalam menentukan banyaknya nilai *k* dapat menggunakan sebuah metode yang disebut *elbow*. Metode *elbow* digunakan untuk menentukan jumlah kelompok yang optimal dalam *k-means clustering*. Metode ini memanfaatkan perubahan dalam variansi dalam setiap kelompok saat jumlah kelompok ditingkatkan (Putu et al., 2017). Metode *elbow* membantu menemukan titik di mana penambahan kelompok baru memberikan manfaat yang lebih kecil dibandingkan dengan penambahan kelompok sebelumnya (Ayu et al., 2019). Oleh karena itu, metode *elbow* dapat membantu dalam menentukan jumlah kelompok yang optimal untuk pengelompokan tempat wisata.

Dengan dilakukannya penelitian ini, diharapkan dapat membantu Dinas Bidang Pariwisata dalam mengetahui jumlah kelompok yang optimal serta karakteristik setiap tempat wisata. Hal ini akan membantu Dinas dalam mengembangkan pengalaman wisata, menyediakan layanan yang lebih sesuai dengan preferensi pengunjung, dan pengembangan destinasi wisata.

1.2 Rumusan Masalah

Bersesuaian dengan masalah yang telah dijabarkan, maka rumusan masalah dalam Tugas Akhir ini ialah bagaimana menerapkan *k-means clustering* untuk mengelompokkan tempat wisata pada Dinas Kebudayaan dan Pariwisata Kabupaten Bojonegoro.

1.3 Batasan Masalah

Batasan masalah dilakukan agar pembahasan masalah tidak terlalu melebar. Berdasarkan pada rumusan masalah yang ada di atas maka batasan masalah dalam menerapkan *data mining* dengan memakai *k-means clustering* terkait pengelompokan tempat wisata yang ada di Kabupaten Bojonegoro yaitu:

1. Penelitian berfokus pada pengelompokan tempat wisata di Kabupaten Bojonegoro guna mengetahui destinasi wisata yang berpotensi untuk dikembangkan sehingga dapat meningkatkan kontribusi sektor wisata.
2. Data wisata yang digunakan dalam penelitian berjumlah 25 tempat wisata.
3. Data yang digunakan dalam proses ini adalah data tahun 2021.
4. Perhitungan *K-Means Clustering* menggunakan tools *Python* dengan *Library Sk-Learn*.
5. Metode yang digunakan untuk menentukan jumlah *cluster* yang optimal adalah metode *elbow*.
6. Atribut yang digunakan dalam pengolahan data terdiri dari Luas Wilayah Destinasi Wisata(%) dan Persentase Jumlah Kunjungan(%).

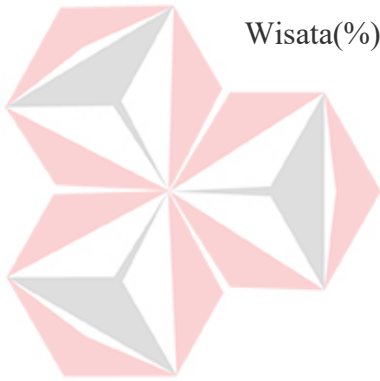
1.4 Tujuan

Mendukung pengelompokan tempat wisata menggunakan metode *k-means clustering* berdasarkan karakteristik masing-masing tempat wisata merupakan tujuan dalam tugas akhir ini.

1.5 Manfaat

Penyusunan tugas akhir ini diharapkan memberikan manfaat, diantaranya yaitu:

1. Membantu mengelompokkan destinasi wisata berdasarkan hasil clustering sehingga dapat diidentifikasi dari sisi 2 atribut yaitu Luas Wilayah Destinasi Wisata(%) dan Persentase Jumlah Kunjungan(%).
2. Menjadi media informasi olah data untuk mengetahui kondisi tempat wisata di Kabupaten Bojonegoro pada tahun 2021 berdasarkan Luas Wilayah Destinasi Wisata(%) dan Persentase Jumlah Kunjungan(%).



UNIVERSITAS
Dinamika

BAB II LANDASAN TEORI

2.1 Penelitian Terdahulu

Penggunaan *Data Mining* dalam proses melakukan penelitian pada dasarnya telah dilakukan oleh penelitian sebelumnya sebagai bahan dalam memperbanyak teori yang diterapkan untuk mengkaji penelitian yang dilakukan oleh penulis. Berdasarkan kajian terdahulu maka dalam hal ini penulis menemukan persamaan dalam judul penelitian namun terdapat juga beberapa perbedaan. Berikut ini adalah penelitian sebelumnya yaitu berupa jurnal mengenai Penelitian yang dilakukan oleh penulis.

Tabel 2.1 Penelitian Terdahulu

Nama Peneliti	Judul Penelitian	Hasil Penelitian
(Savitri et al., 2021)	Pengelempokan Kunjungan Wisata Kabupaten Kulon Progo Tahun 2019 Menggunakan <i>K-Means</i>	Penelitian ini menerapkan metode <i>K-means clustering</i> pada data jumlah pengunjung objek wisata sepanjang tahun 2019. Hasil analisis menunjukkan terdapat tiga cluster objek wisata yaitu C1: jumlah pengunjung rendah sebanyak 20 objek wisata, C2: jumlah pengunjung sedang sebanyak 15 objek wisata, dan C3: jumlah pengunjung tinggi sebanyak 2 objek wisata. Nilai centroid masing masing cluster adalah $C1= 712,9292$, $C2= 6218,5268$, $C3= 30791,9583$. Dengan hasil pengelompokan tersebut, pemangku kepentingan dapat menyusun strategi promosi wisata yang memiliki jumlah pengunjung rendah dengan menyediakan paket wisata dengan destinasi objek wisata pada klaster tersebut.

Terdapat perbedaan penelitian yang akan dilakukan dengan penelitian dari (Savitri et al., 2021) yaitu metode tahapan yang digunakan berupa analisis deskriptif dan hanya satu atribut yang di *cluster* yaitu jumlah kunjungan.

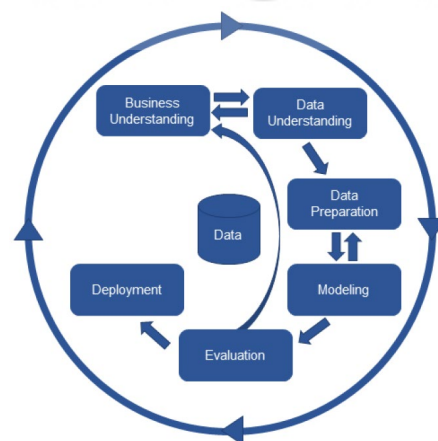
Nama Peneliti	Judul Penelitian	Hasil Penelitian
(Lusianah et al., 2023)	Implementasi Algoritma <i>K-Means</i> Dalam Pengelompokan Jumlah Wisatawan Akomodasi Di Jawa Barat	Penelitian ini memfokuskan bagaimana pengelompokan wisatawan daerah wisata akomodasi dan jenis wisatawan terbanyak yang ramai berkunjung menggunakan algoritma <i>k-means</i> . Teknik pengumpulan data atau pengumpulan data bersumber dari Dinas Pariwisata dan Kebudayaan Provinsi Jawa Barat. Dengan tujuan kiranya menjadi pengetahuan untuk wisatawan serta mampu mempublikasikan wisata di media sosial dan diharapkan akan lebih meningkat jumlah wisatawan yang berkunjung dan meningkatnya pembangunan sarana dan prasarana di setiap daerah akomodas. Hasil dari penelitian ini bahwa daerah akomodasi yang ramai dikunjungi adalah Kota Bandung dengan jenis wisatawan yang datang yaitu pengunjung dalam negeri (Nusantara)
Terdapat perbedaan penelitian yang akan dilakukan dengan penelitian dari (Lusianah et al., 2023) adalah metode yang digunakan <i>KDD (Knowledge Discovery in Database)</i> dan hanya satu atribut yang di <i>cluster</i> yaitu jumlah kunjungan.		
(Pamungkas et al., 2020)	Rekomendasi Wisata Umbul dengan <i>K-Means Clustering</i>	Hasil dari penelitian ini adalah mengelompokkan wisata umbul yang terbagi menjadi 3 cluster yaitu Cukup, Baik dan Terbaik. Hasil pendataan, terdapat 4 wisata umbul yang masuk dalam kluster pertama kategori Indah, yaitu Tirtomoyo, Buto, Pancuran, dan Besuki. Pada klaster ketiga wisata umbul memiliki kategori baik yaitu umbul Tirto Mulyani, Gedaren, Sumber Nila, Manten, Sigedang, dan Kajen. Pada kategori terbaik pada klaster kedua terdapat 8 wisata umbul yaitu umbul Nila, Tirto Mulyono, Ingas,

Nama Peneliti	Judul Penelitian	Hasil Penelitian
		Lumban Tirto, Pongkok, Tirto Raharjo, Jolotundo, dan Tirtomulyono.
Perbedaan pada penelitian yang akan dilakukan dengan penelitian yang telah dilakukan oleh (Pamungkas et al., 2020) adalah atribut yang di <i>cluster</i> lebih berfokus kepada fasilitas yang diberikan seperti kebersihan, ruang ganti, kamar mandi, penjaja makanan, akses jalan, dan keamanan wisata.		

2.2 Data Mining

Suatu proses dalam melakukan pencarian informasi secara otomatis yang dapat bermanfaat dalam menyimpan data yang memiliki ukuran besar merupakan bentuk *data mining* sehingga menemukan pola yang baru dan berguna dalam membuat keputusan (Han et al., 2011). Pada penerapannya, *clustering* salah satu teknik data mining. *Clustering* merupakan teknik yang melakukan peng-*clusteran* data sesuai tingkat kemiripan karakteristik antar data.

Pada penerapan data mining, perlu menjalankan suatu metode agar dapat diterapkan secara sistematis, salah satunya adalah metode *Cross-Standard Industr for Data Mining (CRISP-DM)*. Menurut (Ncr & Clinton, 1999) metode *Cross-Standard Industr for Data Mining (CRISP-DM)* memiliki beberapa tahapan yaitu sebagai berikut:



Gambar 2.1 Metode CRISP-DM

(Sumber: (Han et al., 2011))

1. *Business Understanding*

Hal-hal yang dilaksanakan pada bagian ini meliputi pemahaman mengenai kebutuhan dan juga tujuan berdasarkan perspektif bisnis yang

nantinya akan dapat dimaknai sebagai pengetahuan dalam bentuk pendefinisian masalah yang ada di *data mining* yang nantinya dapat menetapkan rencana dan juga strategi dalam mencapai tujuan data mining.

2. *Data Understanding*

Langkah awal yang dilakukan yaitu mengumpulkan data, diekspresikan data kemudian dievaluasi kualitas data.

3. *Data Preparation*

Pada bagian ini yaitu proses membangun dataset akhir dari beberapa data mentah. Hal-hal yang dilakukan yaitu mengenai proses pembersihan data (*Data Cleaning*), berbagai atribut, record, memilih-milih data (*Data Selection*), serta melakukan transformasi terhadap data tersebut (*Data Transformation*) sehingga dapat diberikan masukan dalam proses pemodelan.

4. *Modelling*

Proses ini akan melibatkan *machine learning* secara langsung yang digunakan untuk menentukan cara data mining, dan juga membantu pembuatan algoritma data mining.

5. *Evaluation*

Tahap ini dilakukan dengan membuat interpretasi dari proses *modelling* yang telah dilakukan.

6. *Deployment*

Pada proses ini perencanaan untuk deployment dimulai selama business understanding dan harus menggabungkan tidak hanya bagaimana untuk menghasilkan nilai model, tetapi juga bagaimana mengkonversi skor keputusan/pengujian, dan bagaimana untuk menggabungkan keputusan dalam sistem operasional.

Namun pada penelitian ini untuk tahapan proses dari *CRISP-DM* hanya dilakukan sampai tahapan *Evaluation* saja karena untuk melanjutkan ke tahapan *Deployment* membutuhkan waktu lebih panjang yang akan dilakukan oleh dinas.

2.3 *K-Means*

Metode ini adalah suatu algoritma dalam proses datamining yang bermanfaat dalam proses pengelompokan data. Terdapat berbagai macam metode untuk

membentuk suatu cluster, diantaranya yaitu dengan membentuk aturan mendiktekan anggota dalam kelompok yang sama sesuai dengan tingkat kesamaan dari anggotanya. Teknik lain yang dapat digunakan yaitu dengan proses pembuatan fungsi untuk mengukur unsur dalam pengelompokan yang menjadi fungsi parameter dalam clustering (H. Witten et al., 2011).

Metode yang dikaitkan dengan algoritma clustering yang memiliki basis jarak dalam membagi data menjadi beberapa cluster dan juga algoritma yang dapat bekerja dalam atribut numerik merupakan makna dari metode K-Means (H. Witten et al., 2011). Sebelum melakukan perhitungan cluster, dalam menyiapkan datanya dilakukan pelabelan data. Pelabelan data adalah bagian penting dari prapemrosesan data yang berfungsi memungkinkan analisis untuk mengisolasi variabel dalam kumpulan data di mana data input dan output diberi label untuk mengklasifikasi data (Han et al., 2011). Kumpulan data berlabel membantu melatih model untuk mengidentifikasi dan memahami pola berulang dalam input yang dimasukkan ke dalamnya agar menghasilkan output yang akurat. Langkah dari perhitungan *K-Means* adalah (Rohmah et al., 2021):

1. Penentuan jumlah *cluster*
2. Penentuan nilai *centroid*
3. Penentuan jarak *centroid*
4. Penentuan posisi *cluster*

Metode yang digunakan untuk mengukur nilai atribut dengan jarak antara centroid adalah metode Euclidean Distance. Rumus dari metode Euclidean Distance adalah sebagai berikut:

$$d(x, y) = \sqrt{\sum_i^n (x_i - y_i)^2}$$

Keterangan :

d : jarak antara x dan y

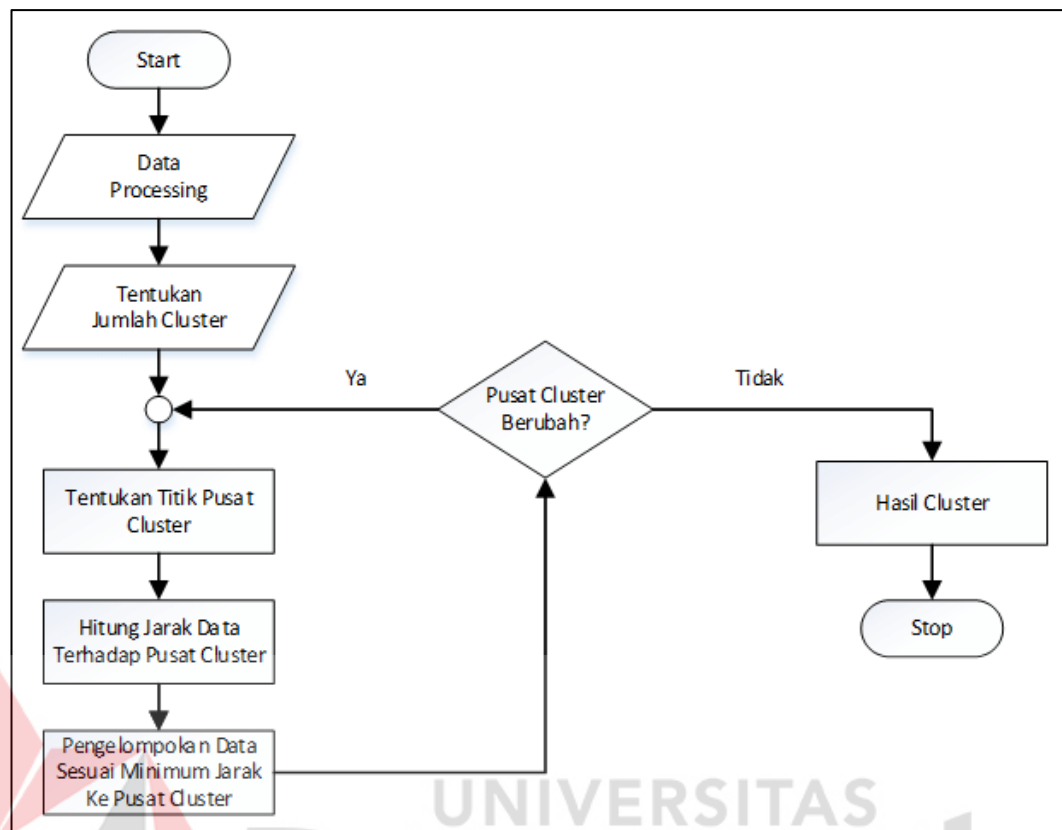
x : nilai *centroid*

y : nilai atribut

x_i : nilai *centroid* ke-i

y_i : nilai atribut ke-i

n : jumlah dimensi



Gambar 2.2 Flowchart Proses *K-Means Clustering*

Sedangkan untuk tahapan dari perhitungan *k-means clustering* menggunakan python adalah sebagai berikut:

1. *Import library (pandas, numpy, sk-learn, matplotlib.pyplot).*
2. *Import csv ke dataset.*
3. *Menentukan titik centroid*
4. *Proses perhitungan k-means clustering*
5. *Visualisasi Clustered Data*
6. *Memasukkan hasil k-means clustering ke dataframe*

2.4 *Phyton*

Bahasa Python adalah bahasa pemrograman tingkat tinggi yang sering digunakan untuk pengembangan perangkat lunak. Diciptakan oleh Guido van Rossum pada tahun 1991, Python dirancang dengan fokus pada keterbacaan kode dan kesederhanaan sintaksisnya. Bahasa ini memiliki filosofi desain yang

mengutamakan kejelasan, yang dikenal dengan prinsip "The Zen of Python" (Romzi & Kurniawan, 2020).

Python memiliki banyak pustaka dan kerangka kerja (framework) yang luas, yang memperluas fungsionalitas bahasa ini dan memudahkan pengembangan aplikasi yang kompleks. Beberapa pustaka terkenal dalam ekosistem Python termasuk NumPy (untuk komputasi numerik), pandas (untuk analisis data), Matplotlib (untuk visualisasi data), Django (untuk pengembangan web), dan TensorFlow (untuk kecerdasan buatan) (Katja & Maufrais, 2010).

Selain itu, Python mendukung pemrograman berorientasi objek, yang memungkinkan pengembang untuk membuat kelas dan objek yang menggabungkan data dan fungsi terkait dalam satu kesatuan. Python juga mendukung pemrograman fungsional dan pemrograman prosedural, sehingga pengembang memiliki fleksibilitas dalam memilih pendekatan yang sesuai dengan kebutuhan proyek.

2.5 Metode *Elbow*

Metode elbow (elbow method) adalah sebuah teknik yang digunakan dalam analisis *clustering* untuk menentukan jumlah cluster yang optimal dalam suatu dataset. Metode ini membantu mengatasi tantangan dalam menentukan jumlah *cluster* yang tepat tanpa mengandalkan penilaian subyektif (Ayu et al., 2019).

Pada dasarnya, metode elbow mengukur varian dalam setiap *cluster* saat jumlah *cluster* bertambah. Grafik yang dihasilkan oleh metode elbow menunjukkan hubungan antara jumlah *cluster* yang digunakan dengan variabilitas dalam *dataset*. Nama "elbow" berasal dari bentuk grafik yang terlihat seperti siku (elbow) pada lengan, apabila nilai cluster pertama dengan nilai cluster kedua memberikan sudut dalam grafik atau nilainya mengalami penurunan paling besar maka jumlah nilai cluster tersebut yang tepat. Untuk mendapatkan perbandingannya adalah dengan menghitung Sum of Square Error (SSE) dari masing-masing nilai cluster. Karena semakin besar jumlah nilai cluster K , maka nilai SSE akan semakin kecil (Putu et al., 2017). Rumus SSE adalah sebagai berikut :

$$SSE = \sum_{K=1}^K \sum_{X_i} |x_i - c_k|^2$$

Keterangan:

K = *cluster ke-c*

X₁ = jarak data obyek ke-*i*

C_k = pusat *cluster ke-i*

Dalam grafik metode *elbow*, jika terdapat titik yang menunjukkan penurunan variabilitas yang signifikan saat jumlah *cluster* meningkat, maka titik tersebut dapat dianggap sebagai "siku" (*elbow*) pada grafik. Jumlah *cluster* yang terkait dengan titik tersebut adalah jumlah *cluster* yang optimal untuk *dataset* yang dianalisis.

Metode *elbow* membantu dalam menentukan jumlah *cluster* yang optimal berdasarkan perubahan variabilitas dalam *dataset*. Dengan menggunakan metode ini, peneliti atau praktisi dapat membuat keputusan yang lebih objektif dalam segmentasi data dan analisis *clustering*, dan menghindari penilaian subyektif yang dapat menghasilkan hasil yang kurang akurat..

2.6 Power BI

Power BI adalah sebuah platform analitik bisnis yang dikembangkan oleh *Microsoft*. Platform ini menyediakan berbagai alat dan fitur untuk mengumpulkan, menganalisis, dan memvisualisasikan data dari berbagai sumber yang berbeda. *Power BI* memungkinkan pengguna untuk membuat *dashboard* interaktif, laporan, dan visualisasi data yang dapat digunakan untuk mendapatkan wawasan dan pemahaman yang lebih dalam tentang kinerja bisnis (Gowthami & Pavan Kumar, 2017).

Salah satu fitur utama *Power BI* adalah kemampuannya untuk menggabungkan data dari berbagai sumber yang berbeda, termasuk basis data, file *excel*, layanan *cloud*, dan banyak lagi. Pengguna dapat mengimpor data ke *Power BI*, membersihkan dan mengolah data, serta menghubungkan data dengan cara yang intuitif dan mudah digunakan (Edwin Bororing, 2022).

Setelah data diimpor, *Power BI* menyediakan berbagai jenis visualisasi data yang dapat disesuaikan, seperti diagram batang, grafik lingkaran, peta, dan

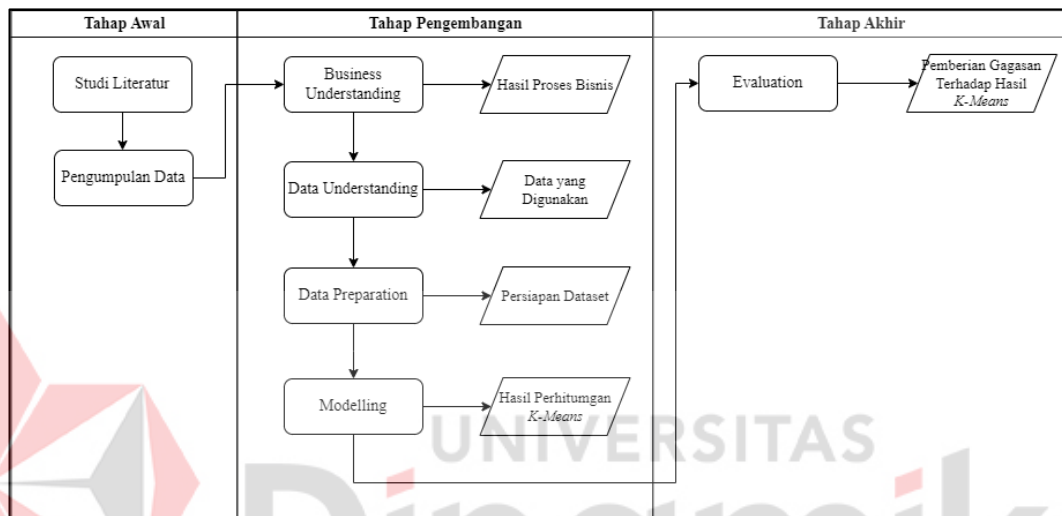
sebagainya. Pengguna dapat mengatur visualisasi ini dalam *dashboard* yang interaktif, sehingga memungkinkan pengguna untuk menjelajahi dan menganalisis data dengan cepat dan efisien. *Power BI* juga menyediakan fitur pembuatan laporan yang fleksibel, yang memungkinkan pengguna untuk membuat laporan yang rapi dan informatif.



UNIVERSITAS
Dinamika

BAB III METODOLOGI PENELITIAN

Penerapan Data Mining dalam kajian ini terbagi ke dalam tiga proses yaitu proses awal, proses pengembangan dan proses akhir. Gambaran tahapan penelitian dapat pada gambar 3.1.



Gambar 3.1 Metodologi Penelitian

3.1 Tahap Awal

Pada tahapan awal akan melakukan kegiatan yang bertujuan untuk dapat menemukan perumusan masalah yang dapat diteliti lebih detail. Kegiatan yang dapat dilakukan pada tahapan ini dapat berupa melakukan studi literatur dan pengumpulan data.

3.1.1 Studi Literatur

Pada proses studi literatur yang dilakukan oleh penulis, penulis melaksanakan dan melakukan pengkajian yang bersesuaian dengan topik yang telah diambil oleh penulis. Selain itu, untuk mendukung topik yang telah diambil maka penulis melakukan pencarian referensi terkait teori yang bermanfaat dalam proses penyelesaian masalah pada penelitian. Maksud dari referensi penelitian ini yaitu:

1. Penelitian sebelumnya
2. *Data Mining*

3. *K-Means Clustering*

Pada penelitian ini, referensi yang digunakan sesuai dengan yang telah disebutkan di atas yaitu diperoleh melalui buku, jurnal, beberapa artikel laporan penelitian sebelumnya, dan juga menggunakan situs internet. Manfaat referensi ini yaitu dijadikan sebagai dasar dan juga panduan teori dalam proses pelaksanaan studi yang bermanfaat untuk menguatkan berbagai macam masalah yang telah dijelaskan pada tahap sebelumnya.

3.1.2 Pengumpulan Data

Pada proses pengumpulan data, langkah yang dilakukan penulis yaitu dengan melaksanakan wawancara kepada berbagai pihak yang memiliki kaitan dalam proses penelitian ini sebagai upaya untuk memperoleh informasi terkait dengan keadaan dan juga proses pengelolaan yang sedang berjalan. Pihak yang dimaksud dalam hal ini adalah pada Bidang Pariwisata Dinas Kebudayaan dan Pariwisata Kabupaten Bojonegoro. Wawancara dilakukan dengan Kepala Bidang Pariwisata untuk mengetahui pengolahan data pada bidang pariwisata dan indikator kinerjanya.

3.2 Tahap Pengembangan

Pada proses pengembangan, terdapat beberapa proses atau langkah di dalam penelitian ini, setiap langkah dibantu dengan pemanfaatan. Pada penelitian ini terus yang dimanfaatkan yaitu *Python dan Power BI*.

3.2.1 *Business Understanding*

Dalam proses ini, penulis melaksanakan wawancara kembali dengan Kepala Bidang Pariwisata untuk melakukan analisis dan penjabaran masalah yang sedang dihadapi terutama dalam hal pengelompokan tempat wisata guna pengembangan pariwisata berdasarkan data-data yang telah diperoleh sebelumnya.

3.2.2 *Data Understanding*

Pada tahap ini, peneliti menghimpun data yang relevan dengan objek penelitian. Data yang dibutuhkan dapat dilihat pada tabel di bawah ini.

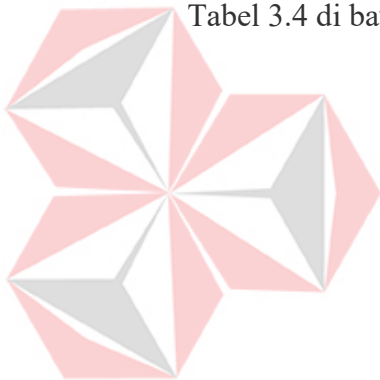
Tabel 3.1 Data yang Dibutuhkan

DATA	INFORMASI YANG DIBUTUHKAN PENELITI
Data Kunjungan Wisatawan	<ol style="list-style-type: none"> 1. Mengetahui jumlah wisatawan yang berkunjung ke tempat wisata 2. Mengetahui kapasitas pengunjung tempat wisata
Data Luas Wilayah Pariwisata	<ol style="list-style-type: none"> 1. Mengetahui luas wilayah tempat wisata 2. Mengetahui luas wilayah tempat wisata yang sudah digunakan

3.2.3 Data Preparation

Dalam tahap ini dimana menyiapkan data yang akan digunakan ke tahap *clustering* dengan memperhatikan atribut yang terdiri dari nama tempat wisata, persentase luas wilayah destinasi wisata (diperoleh dari hasil perbandingan luas wilayah keseluruhan dan wilayah yang digunakan) dan persentase jumlah kunjungan (diperoleh dari perbandingan rata-rata kunjungan dengan kapasitas maksimal tempat wisata). Data tersebut dapat dilihat pada Tabel 3.2, Tabel 3.3, dan

Tabel 3.4 di bawah ini.



Tabel 3.2 Data Jumlah Kunjungan Wisata

No.	Nama Wisata	Januari	Februari	Maret	April	Mei	Juni	Juli	Agustus	September	Oktober	November	Desember
1	Dander Waterpark	11697	3403	3504	5616	1592	1634	4226	2944	4419	4543	5637	12175
2	Kayangan Api	9826	2372	4220	3280	6630	4748	2047	1938	3677	3273	4016	4928
3	Waduk Pacal	3135	1306	3884	2724	5009	4550	2070	1180	3270	3042	3688	4208
4	Taman Pinggir Gawan (TPG)	17667	6638	2660	7797	1415	2546	1150	3946	1250	2100	1700	1925
5	Negeri Atas Angin	1589	444	438	438	408	2626	701	287	870	950	658	649
6	Agrowisata Belimbing	22926	10925	7800	6200	8500	8800	6300	7900	8200	8900	5900	9634
7	Agrowisata Jambu Padang	1600	2500	1500	2460	3390	3845	3590	1800	4100	2880	2300	3300
8	Wisata Edukasi Gerabah	2051	2723	5883	4367	676	622	203	446	704	3974	2135	5529
9	Wali Kidangan	352	372	412	458	380	356	412	426	784	698	584	978
10	Teksas Wonocolo	108	186	300	250	192	300	125	106	160	175	338	172
...

(Sumber: Disbudpar Kabupaten Bojonegoro)

Tabel 3.2 berisi jumlah kunjungan dari masing-masing tempat wisata dalam setahun yang nantinya digunakan untuk mengukur rata-rata kunjungan dalam setahun.

Tampilan data secara keseluruhan dapat dilihat pada lampiran 3.

Tabel 3.3 Data Kapasitas Kunjungan Tempat Wisata

No.	Nama Wisata	Kapasitas Kunjungan
1	Dander Waterpark	8000
2	Kayangan Api	7500
3	Waduk Pacal	4000
4	Taman Pinggir Gawan (TPG)	10000
5	Negeri Atas Angin	1000
6	Agrowisata Belimbing	15000
7	Agrowisata Jambu Padang	10000
8	Wisata Edukasi Gerabah	6000
9	Wali Kidangan	1000
10	Teksas Wonocolo	300
11	Rumah Singgah	200
12	Sendang Tirta Arum	500
13	Waduk Bendo	2500
14	Penangkaran Rusa Malo	2500
15	Growgoland	1500
16	Rumah Tua Padangan	500
17	Mumbul Kalianyar	1000
18	Puthuk Kreweng	1500
19	Banyu Kuning	3000
20	Outbond Toyoaji	1500
21	Masyarakat Samin	2000
22	Agrowisata Salak Wedi	10000
23	Wisata Edukasi Pejambon	1000
24	Petilasan Angling Dharma	800
25	Sumber Maha Dewi	5000

Selanjutnya yang berisikan data terkait nama tempat wisata dan kapasitas kunjungan tempat wisata dapat dilihat pada tabel 3.3.

Tabel 3.4 Data Luas Wilayah Tempat Wisata

No.	Nama Wisata	Luas Wilayah (ha)	Luas yang Sudah Digunakan (ha)
1	Dander Waterpark	1.8	1.5
2	Kayangan Api	16	4
3	Waduk Pacal	387.8	312
4	Taman Pinggir Gawan (TPG)	8	1.5
5	Negeri Atas Angin	9.84	7.69
6	Agrowisata Belimbing	20.4	6.8
7	Agrowisata Jambu Padang	7	7
8	Wisata Edukasi Gerabah	0.14	0.14
9	Wali Kidangan	1.6	1.6

10	Teksas Wonocolo	10	8
11	Rumah Singgah	0.0015	0.0015
12	Sendang Tirta Arum	0.8	0.8
13	Waduk Bendo	3	1
14	Penangkaran Rusa Malo	0.48	0.28
15	Growgoland	6	3
16	Rumah Tua Padangan	0.005	0.005
17	Mumbul Kalianyar	1.1	1.1
18	Puthuk Kreweng	7.17	7.17
19	Banyu Kuning	3	2.3
20	Outbond Toyoaji	5	3.8
21	Masyarakat Samin	74.7	74.7
22	Agrowisata Salak Wedi	0.8	0.8
23	Wisata Edukasi Pejambon	190	164
24	Petilasan Angling Dharma	0.5	0.5
25	Sumber Maha Dewi	1.2	1

(Sumber: Disbudpar Kabupaten Bojonegoro)

Tabel 3.4 berisikan data terkait nama tempat wisata, luas wilayah tempat wisata secara keseluruhan, dan luas tempat wisata yang digunakan

3.2.4 Modelling

Dalam pelaksanaan pengelompokkan menggunakan *k-means clustering*, maka teknik yang dimanfaatkan yaitu dengan bantuan *tool Python*. Proses persiapan dataset dalam menghitung *k-means* dapat dilihat dalam Tabel 3.6.

Tabel 3. 5 Contoh Dataset

No.	Nama Tempat Wisata	Persentase Jumlah Kunjungan(%)	Luas Wilayah Destinasi Wisata (%)
		Diperoleh dari perbandingan rata-rata kunjungan dengan kapasitas maksimal tempat wisata	Diperoleh dari perbandingan antara luas wilayah yang digunakan dengan wilayah secara keseluruhan
1	Dander Waterpark	0.639	0.83
2	Kayangan Api	0.566	0.25
3	Waduk Pacal	0.793	0.8
4	Taman Pinggir Gawan (TPG)	0.423	0.19
5	Negeri Atas Angin	0.838	0.78

No.	Nama Tempat Wisata	Persentase Jumlah Kunjungan(%)	Luas Wilayah Destinasi Wisata (%)
		Diperoleh dari perbandingan rata-rata kunjungan dengan kapasitas maksimal tempat wisata	Diperoleh dari perbandingan antara luas wilayah yang digunakan dengan wilayah secara keseleruhan
6	Agrowisata Belimbing	0.622	0.33
7	Agrowisata Jambu Padang	0.277	1.00
8	Wisata Edukasi Gerabah	0.407	1.00
9	Wali Kidangan	0.518	1.00
10	Dander Waterpark	0.639	0.83

Contoh langkah-langkah sederhana perhitungan *K-Means Clustering* adalah sebagai berikut:

1. Tentukan jumlah *cluster*
2. Pilih *centroid* awal secara acak sesuai jumlah *cluster* yang telah ditentukan

Tabel 3.6 Titik *Centroid* Awal

Titik <i>Centroid</i> Awal	Persentase Jumlah Kunjungan(%)	Luas Wilayah Destinasi Wisata (%)
<i>Cluster 1</i>	0.566	0.25
<i>Cluster 2</i>	0.622	0.33
<i>Cluster 3</i>	0.407	1.00

3. Hitung jarak data ke *centroid* dengan rumus Euclidean Distance seperti berikut ini:

$$\begin{aligned}
 d_{11} &= \sqrt{(0.639 - 0.566)^2 + (0.83 - 0.25)^2} \\
 &= \sqrt{(0.073)^2 + (0.58)^2} \\
 &= \sqrt{0.341729} \\
 &= 0.584
 \end{aligned}$$

Dari perhitungan di atas dapat diketahui bahwa jarak data pertama dengan cluster pertama adalah 0.584.

$$\begin{aligned}
 d_{12} &= \sqrt{(0.639 - 0.622)^2 + (0.83 - 0.33)^2} \\
 &= \sqrt{(0.017)^2 + (0.5)^2} \\
 &= \sqrt{0.250289}
 \end{aligned}$$

$$= 0.5002$$

Dari perhitungan di atas dapat diketahui bahwa jarak data pertama dengan cluster kedua adalah 0.5002.

$$\begin{aligned} d_{13} &= \sqrt{(0.639 - 0.407)^2 + (0.83 - 1.00)^2} \\ &= \sqrt{(0.232)^2 + (-0.17)^2} \\ &= \sqrt{0.082724} \\ &= 0.287 \end{aligned}$$

Dari perhitungan di atas dapat diketahui bahwa jarak data pertama dengan cluster kedua adalah 0.287.

Acuan ditentukannya masuk dalam *cluster 1*, *cluster 2*, atau *cluster 3* didapatkan sesuai dengan jarak minimum data dengan titik centroid yang sudah ditentukan. Nantinya dari masing-masing cluster tersebut akan dianalisis karakteristiknya dan dikaitkan sesuai dengan parameter clustering pada tahap selanjutnya setelah semua data melalui semua tahapan perhitungan *k-means*.

Tabel 3.7 Contoh Hasil Perhitungan Tiap Cluster

Persentase Jumlah Kunjungan(%)	Luas Wilayah Destinasi Wisata (%)	Jarak			Cluster
		C1	C2	C3	
0.639	0.83	0.584	0.5002	0.287	3
0.566	0.25	0	0.2094	0.6115	1
0.793	0.8	0.5295	0.2449	0.316	2
0.423	0.19	0.1466	0.3656	0.7501	1
0.838	0.78	0.5529	0.2715	0.3694	3
0.622	0.33	0.0624	0	0.5539	2
0.277	1.00	0.8515	0.9409	0.24	3
0.407	1.00	0.7215	0.8109	0	3
0.518	1.00	0.6105	0.6999	0.001	3
0.639	0.83	0.4065	0.3199	0.193	3

Tabel 3.9 merupakan hasil perhitungan dari jarak masing-masing data dengan cluster yang sudah ditentukan dan pengelompokan cluster. Hasil cluster tersebut masih bisa berubah karena perhitungan baru melalui satu iterasi. Untuk mendapatkan hasil cluster yang stabil perlu dilakukan dengan melalui beberapa iterasi dengan batas yang tidak ditentukan hingga data benar-benar sudah stabil atau tidak berpindah cluster lagi. Iterasi dilakukan hingga didapatkan matrik jarak. Jarak yang telah dihitung sebelumnya dibandingkan dan dipilih yang terdekat dalam suatu kelompok.

4. Menghitung nilai rata-rata tiap *cluster* dengan memperbaharui *centroid*

Tabel 3.8 Contoh Titik *Cluster* Baru

Titik <i>Centroid</i> Baru	Persentase Jumlah Kunjungan(%)	Luas Wilayah Destinasi Wisata (%)
<i>Cluster 1</i>	0.470	0.338
<i>Cluster 2</i>	0.814	0.657
<i>Cluster 3</i>	0.554	0.955

5. Ulangi Melakukan pengulangan proses ke-3 jika masih terdapat perpindahan kelompok.

Pada penelitian ini, proses perhitungan dilakukan dengan menggunakan tool Python. Terdapat beberapa library Python yang dimanfaatkan dalam penghitungan pengelompokan dengan k-means clustering yaitu:

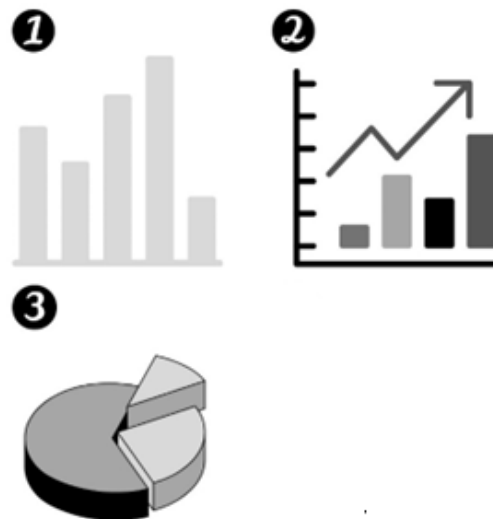
- a) *NumPy* adalah library Python yang berguna untuk komputasi numerik, memanipulasi dan mengoperasikan array
- b) *Pandas* memiliki fungsi mengolah sumber data yang dihubungkan dengan python.
- c) *Scikit-learn* memiliki fungsi perhitungan *k-means* dan evaluasi dengan metode *elbow*.
- d) *Matplotlib.pyplot* berfungsi untuk visualisasi data.

Penggunaan *python* dalam proses *clustering* dapat dilihat pada lampiran 1.

3.3 Tahap Akhir

3.3.1 *Evaluation*

Pada tahapan evaluasi ini, proses yang dilaksanakan yaitu memberikan gagasan terhadap hasil yang telah diperoleh dari data mining yang didapatkan pada pembuatan model dalam proses sebelumnya. Tahapan evaluasi yang dilaksanakan yaitu melakukan uji coba untuk menentukan jumlah *cluster* yang optimal dalam penelitian ini. Setelah itu, keseluruhan data yang telah diolah akan dibuat sebuah *user interface* menggunakan *Power BI*. Pembuatan *dashboard* dilakukan dengan menghubungkan antara data Bidang Pariwisata yang telah jadi dengan *Power BI*.



Gambar 3.2 Contoh Diagram Dashboard

Berikut adalah beberapa contoh bentuk diagram dalam *dashboard* beserta kegunaannya yang akan dibuat sesuai pada Gambar 3.2:

1. *Bar Chart*: diagram batang digunakan untuk membandingkan kuantitas atau nilai di antara beberapa kategori sehingga dapat membandingkan jumlah luas wisata dari masing-masing tempat wisata.
2. *Line Chart*: diagram garis digunakan untuk menunjukkan perubahan nilai seiring waktu. *Line chart* berguna untuk memantau tren jumlah kunjungan dalam jangka waktu tertentu.
3. *Pie Chart*: diagram lingkaran digunakan untuk membandingkan proporsi atau persentase dari beberapa kategori. *Pie chart* dapat membantu bidang pariwisata dalam melihat hasil clustering destinasi wisata.

BAB IV

HASIL DAN PEMBAHASAN

Pada hasil dan pembahasan akan berisikan tentang proses yang dilakukan dari cara pengerjaan yang sebelumnya telah disusun pada bab sebelumnya. Proses penyusunan pada bab ini diawali dari melakukan tahap awal kemudian dilanjutkan pada tahap pengembangan dan diakhiri pada tahap akhir.

4.1 Data Preparation

Hasil pengolahan data yang dilakukan pada data internal dengan menggunakan *microsoft excel* yaitu sebagai berikut:

4.1.1 Hasil Pengolahan Data Kunjungan Wisata

Data kunjungan wisata diolah untuk mendapatkan data rata-rata kunjungan wisata dalam setahun. Hasil olah data rata-rata kunjungan diperoleh dari menghitung rata-rata kunjungan mulai bulan Januari hingga bulan Desember, sedangkan untuk mendapatkan persentase kunjungan wisata diperoleh dari menghitung perbandingan rata-rata kunjungan dengan kapasitas maksimal tempat wisata. Hasil data dilihat pada Tabel 4.1 dan dapat dilihat secara lengkap pada lampiran 3.

Sedangkan untuk hasil dari persentase jumlah kunjungan dapat dilihat pada Tabel 4.2

Tabel 4.2 Persentase Jumlah Kunjungan

Nama Wisata	Rata-Rata Kunjungan	Kapasitas Tempat Wisata	Persentase Jumlah Kunjungan
Dander Waterpark	5116	8000	64%
Kayangan Api	4246	7500	57%
Waduk Pacal	3172	4000	79%
Taman Pinggir Gawan (TPG)	4233	10000	42%
Negeri Atas Angin	838	1000	84%
Agrowisata Belimbing	9332	15000	62%
Agrowisata Jambu Padang	2772	10000	28%
Wisata Edukasi Gerabah	2443	6000	41%
Wali Kidangan	518	1000	52%
Teksas Wonocolo	201	300	67%
Rumah Singgah	164	200	82%
Sendang Tirta Arum	259	500	52%
Waduk Bendo	1924	2500	77%
Penangkaran Rusa Malo	667	2500	27%
Growgoland	1512	1500	100%
Rumah Tua Padangan	128	500	26%
Mumbul Kalianyar	867	1000	87%
Puthuk Kreweng	938	1500	63%
Banyu Kuning	2231	3000	74%
Outbond Toyoaji	1107	1500	74%
Masyarakat Samin	1203	2000	60%
Agrowisata Salak Wedi	1039	10000	10%
Wisata Edukasi Pejambon	766	1000	77%
Petilasan Angling Dharma	414	800	52%
Sumber Maha Dewi	3644	5000	73%

4.1.2 Hasil Pengolahan Data Luas Wilayah Pariwisata

Data luas wilayah pariwisata diolah untuk mendapatkan perbandingan antara luas wilayah yang digunakan dengan wilayah secara keseluruhan yang diubah ke dalam skala persentase. Hasil data dapat dilihat pada Tabel 4.3.

Tabel 4.3 Data Luas Wilayah Pariwisata

No.	Nama Wisata	Total Luas Wilayah (ha)	Luas Digunakan (ha)	Luas potensi Wilayah Digunakan (%)
1	Dander Waterpark	1.8	1.31	73%

No.	Nama Wisata	Total Luas Wilayah (ha)	Luas Digunakan (ha)	Luas potensi Wilayah Digunakan (%)
2	Kayangan Api	16	12	75%
3	Waduk Pacal	387.8	267.03	69%
4	Taman Pinggir Gawan (TPG)	8	6.32	19%
5	Negeri Atas Angin	9.84	7.69	78%
6	Agrowisata Belimbing	20.4	6.8	33%
7	Agrowisata Jambu Padang	7	4.69	100%
8	Wisata Edukasi Gerabah	0.14	0.14	100%
9	Wali Kidangan	1.6	1.6	100%
10	Teksas Wonocolo	10	8	80%
11	Rumah Singgah	0.0015	0.0015	100%
12	Sendang Tirta Arum	0.8	0.8	100%
13	Waduk Bendo	3	1	33%
14	Penangkaran Rusa Malo	0.48	0.28	58%
15	Growgoland	6	3	50%
16	Rumah Tua Padangan	0.005	0.005	100%
17	Mumbul Kalianyar	1.1	1.1	100%
18	Puthuk Kreweng	7.17	7.17	100%
19	Banyu Kuning	3	2.3	77%
20	Outbond Toyoaji	5	3.8	76%
21	Masyarakat Samin	74.7	74.7	100%
22	Agrowisata Salak Wedi	0.8	0.8	100%
23	Wisata Edukasi Pejambon	190	164	86%
24	Petilasan Angling Dharma	0.5	0.5	100%
25	Sumber Maha Dewi	1.2	1	83%

4.2 Modelling

Penelitian ini dilakukan dengan mengambil data dari tahun 2021 yang terdiri dari 25 tempat wisata dengan atribut yang digunakan yaitu Nama Tempat Wisata, Persentase Jumlah Kunjungan (%), dan Luas Wilayah Destinasi Wisata (%). Jumlah cluster yang akan digunakan adalah $K=2$ sampai dengan $K=10$. Proses implementasi data mining diolah dengan menggunakan *python*. Proses perhitungan *k-means clustering* dapat dilihat pada tahapan berikut ini.

1. Proses mengimport *library*

```
import numpy as np
import pandas as pd
from sklearn.cluster import KMeans
```

NumPy digunakan untuk proses manipulasi data dan mengoperasikan *array*, *pandas* digunakan untuk proses membaca *dataset* yang akan diolah, sedangkan *sklearn.cluster* digunakan untuk proses *k-means clustering*.

2. Proses insert *dataset* berformat *csv* yang akan diolah

```
dataset = pd.read_csv("wisata_update.csv")
data = dataset.iloc[:, 1:5]
data.head(25)
```

Tabel 4.4 Hasil *Insert Dataset*

Nama Wisata	Persentase Jumlah Kunjungan	Persentase Luas Wilayah
Dander_Waterpark	0.639	0.83
Kayangan_Api	0.566	0.25
Waduk_Pacal	0.793	0.80
Taman_Pinggir_Gawan	0.423	0.19
Negeri_Atas_Angin	0.838	0.78
...

3. Proses menentukan titik *centroid*

```
centroids = kmeans.cluster_centers_
print("Centroid:")
for centroid in centroids:
    print(centroid)
```

```
Centroid:
[0.407 1.00]
[0.423 0.19]
[0.739625 0.80375]
[1. 0.5]
[0.21233333 1.00]
[0.267 0.58]
[0.65233333 0.30333333]
[0.556 1.00]
[0.8425 1.00]
```

4. Proses inialisasi *i* (jumlah cluster antara 2-10), membuat objek *k-means* dengan jumlah kelompok yang diinginkan, serta melakukan proses clustering data

```
for i in range(2, 10):
    kmeans = KMeans(n_clusters = i, init = 'k-means++', random_state = 42)
    kmeans.fit(data)

    k = i
    kmeans = KMeans(n_clusters=k)
    y_predicted = kmeans.fit(df)
```

5. Proses mendapatkan label kelompok dari setiap data

```
labels = kmeans.labels_
df['Cluster']= labels
df
```

Tabel 4.5 Hasil Pelabelan Setelah Proses *Clustering*

Nama Wisata	Persentase Jumlah Kunjungan	Persentase Luas Wilayah	Cluster
Dander_Waterpark	0.639	0.83	2
Kayangan_Api	0.566	0.25	6
Waduk_Pacal	0.793	0.80	2
Taman_Pinggir_Gawan	0.423	0.19	1
Negeri_Atas_Angin	0.838	0.78	2
...

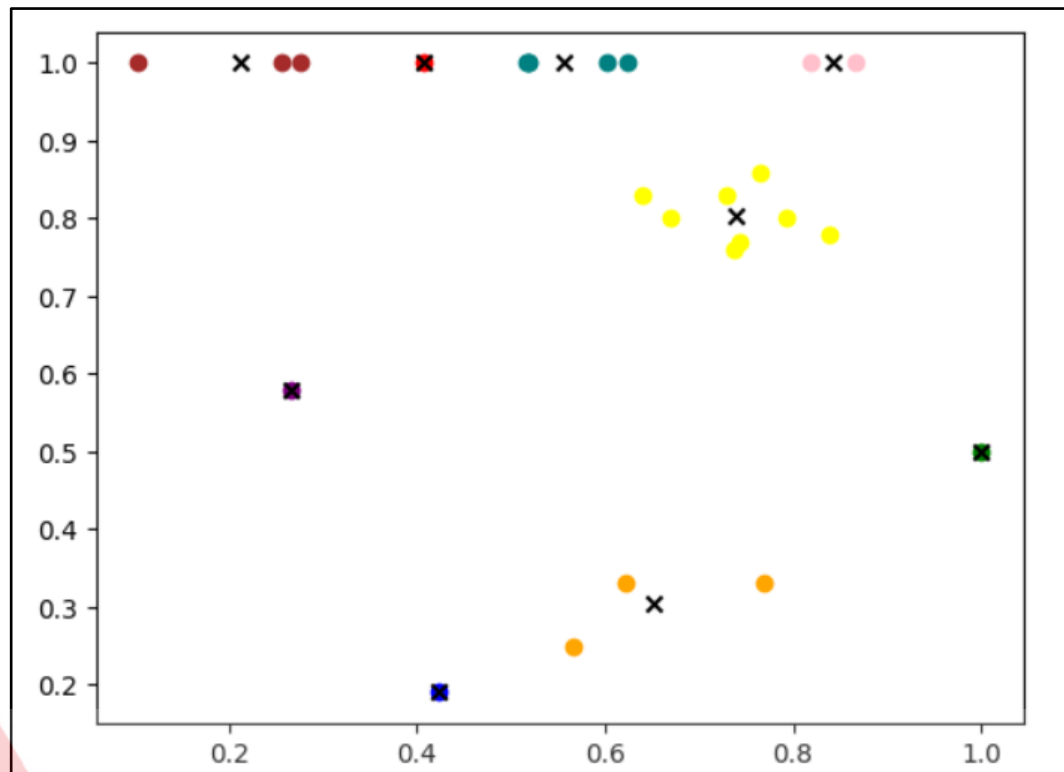
6. Proses visualisi *scatter* menggunakan library *matplotlib.pyplot* untuk melihat persebaran data

```
import matplotlib.pyplot as plt

dat0 = df[(df["Cluster"] == 0)]
dat1 = df[(df["Cluster"] == 1)]
dat2 = df[(df["Cluster"] == 2)]
dat3 = df[(df["Cluster"] == 3)]
dat4 = df[(df["Cluster"] == 4)]
dat5 = df[(df["Cluster"] == 5)]
dat6 = df[(df["Cluster"] == 6)]
dat7 = df[(df["Cluster"] == 7)]
dat8 = df[(df["Cluster"] == 8)]
dat9 = df[(df["Cluster"] == 9)]

plt.scatter(dat0["Persentase Jumlah Kunjungan"], dat0["Persentase Luas Wilayah"], color='red')
plt.scatter(dat1["Persentase Jumlah Kunjungan"], dat1["Persentase Luas Wilayah"], color='blue')
plt.scatter(dat2["Persentase Jumlah Kunjungan"], dat2["Persentase Luas Wilayah"], color='yellow')
plt.scatter(dat3["Persentase Jumlah Kunjungan"], dat3["Persentase Luas Wilayah"], color='green')
plt.scatter(dat4["Persentase Jumlah Kunjungan"], dat4["Persentase Luas Wilayah"], color='brown')
plt.scatter(dat5["Persentase Jumlah Kunjungan"], dat5["Persentase Luas Wilayah"], color='purple')
plt.scatter(dat6["Persentase Jumlah Kunjungan"], dat6["Persentase Luas Wilayah"], color='orange')
plt.scatter(dat7["Persentase Jumlah Kunjungan"], dat7["Persentase Luas Wilayah"], color='teal')
plt.scatter(dat8["Persentase Jumlah Kunjungan"], dat8["Persentase Luas Wilayah"], color='pink')
plt.scatter(dat9["Persentase Jumlah Kunjungan"], dat9["Persentase Luas Wilayah"], color='gray')

plt.scatter(centroids[:,0], centroids[:,1], marker="x", color='black')
plt.show()
```

Gambar 4.1 Hasil *Scatter*7. Proses menghitung banyaknya jumlah anggota dari setiap *cluster*

```
dataset["Cluster"]=labels
df["Cluster"].value_counts()
```

Tabel 4.6 Jumlah Anggota Dari Setiap *Cluster*

Cluster	Jumlah
0	1
1	1
2	8
3	1
4	3
5	1
6	3
7	5
8	2

Setelah melalui proses perhitungan *k-means clustering* dengan inisialisasi $K=2$ sampai $K=10$ didapatkan total 9 *cluster*. Selanjutnya, pada tahap *evaluation* akan dilakukan sebuah evaluasi jumlah *cluster* yang optimal agar hasil dari *clustering* tempat wisata lebih akurat.

4.3 Evaluation

Pada tahap *evaluation* hasil dari perhitungan *clustering* akan dievaluasi dengan menggunakan jumlah *cluster* yang optimal. Data yang diuji mengambil data dari tahun 2021 yang terdiri dari 25 tempat wisata dengan atribut yang digunakan yaitu Nama Tempat Wisata, Persentase Jumlah Kunjungan (%) dan Persentase Luas Wilayah (%). Jumlah *cluster* yang akan diuji adalah $K=2$ sampai dengan $K=10$. Untuk melakukan coba dalam menentukan jumlah K terbaik, maka uji coba akan dilakukan dalam 3 kali uji coba dengan jumlah data secara acak yaitu 10 data, 15 data, dan 25 data.

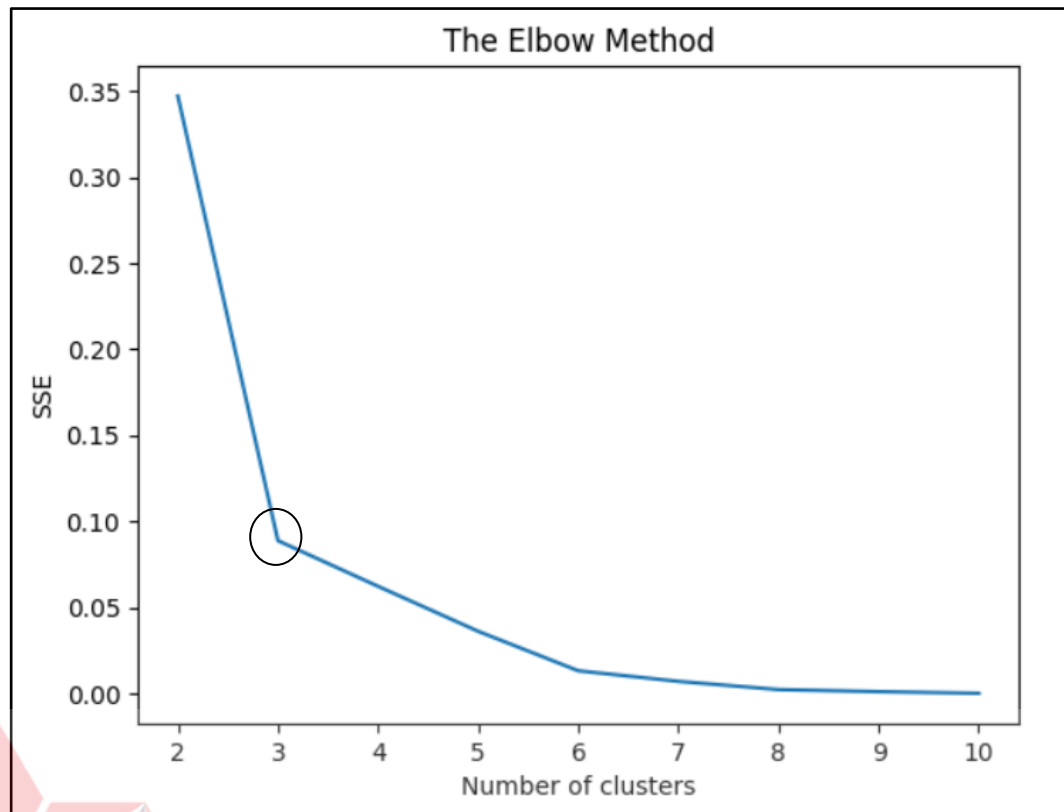
4.3.1 Uji Coba 1 dengan 10 Data Tempat Wisata

Dari hasil proses perhitungan *Sum of Square Error* terhadap 10 data tempat wisata maka hasil yang mengalami penurunan yang paling besar adalah pada $K=3$. Hasil dari perhitungan *Sum of Square Error* 10 data dapat dilihat pada Tabel 4.8.

Tabel 4.7 Hasil Perhitungan *Sum of Square Error* 10 Data

<i>Cluster</i>	Hasil <i>Sum Square of Error</i>	Selisih
2	0,3472	0,3472
3	0,0887	0,2585
4	0,0621	0,0266
5	0,0360	0,0261
6	0,0130	0,023
7	0,0069	0,0061
8	0,0021	0,0048
9	0,0009	0,0012
10	0	0,0009

Hasil dari grafik *Sum of Square Error* 10 data dapat dilihat pada Gambar 4.2.



Gambar 4.2 Grafik *Sum of Square Error* 10 Data

Dari hasil proses perhitungan *Sum of Square Error* terhadap 10 data tempat wisata maka hasil grafik yang mengalami penurunan yang paling besar adalah pada $K=3$.

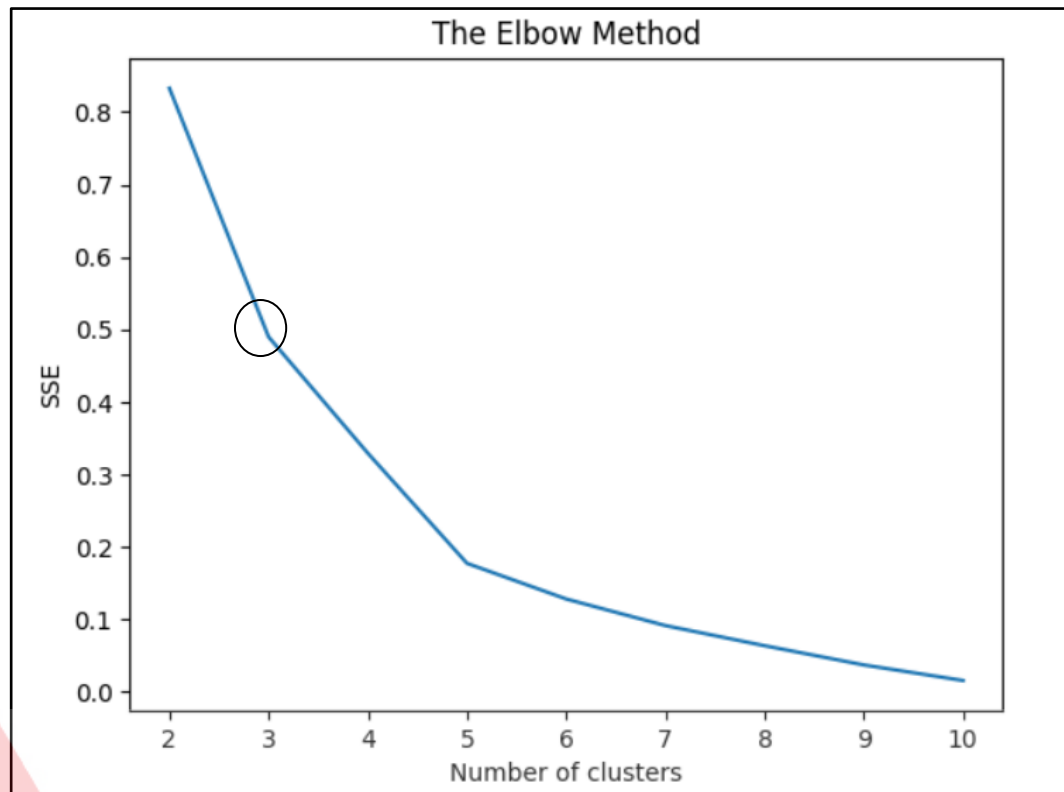
4.3.2 Uji Coba 2 dengan 15 Data Tempat Wisata

Dari hasil proses perhitungan *Sum of Square Error* terhadap 15 data tempat wisata maka hasil yang mengalami penurunan yang paling besar adalah pada $K=3$. Hasil dari perhitungan *Sum of Square Error* 15 data dapat dilihat pada Tabel 4.9.

Tabel 4.8 Hasil Perhitungan *Sum of Square Error* 15 Data

Cluster	Hasil <i>Sum Square of Error</i>	Selisih
2	0,8329	0,8329
3	0,4895	0,3434
4	0,3290	0,1605
5	0,1768	0,1522
6	0,1276	0,0492
7	0,0909	0,0367
8	0,0631	0,0278
9	0,0366	0,0265
10	0,0151	0,0215

Hasil dari grafik *Sum of Square Error* 15 data dapat dilihat pada Gambar 4.3.



Gambar 4.3 Grafik *Sum of Square Error* 15 Data

Dari hasil proses perhitungan *Sum of Square Error* terhadap 15 data tempat wisata maka hasil grafik yang mengalami penurunan yang paling besar adalah pada $K=3$.

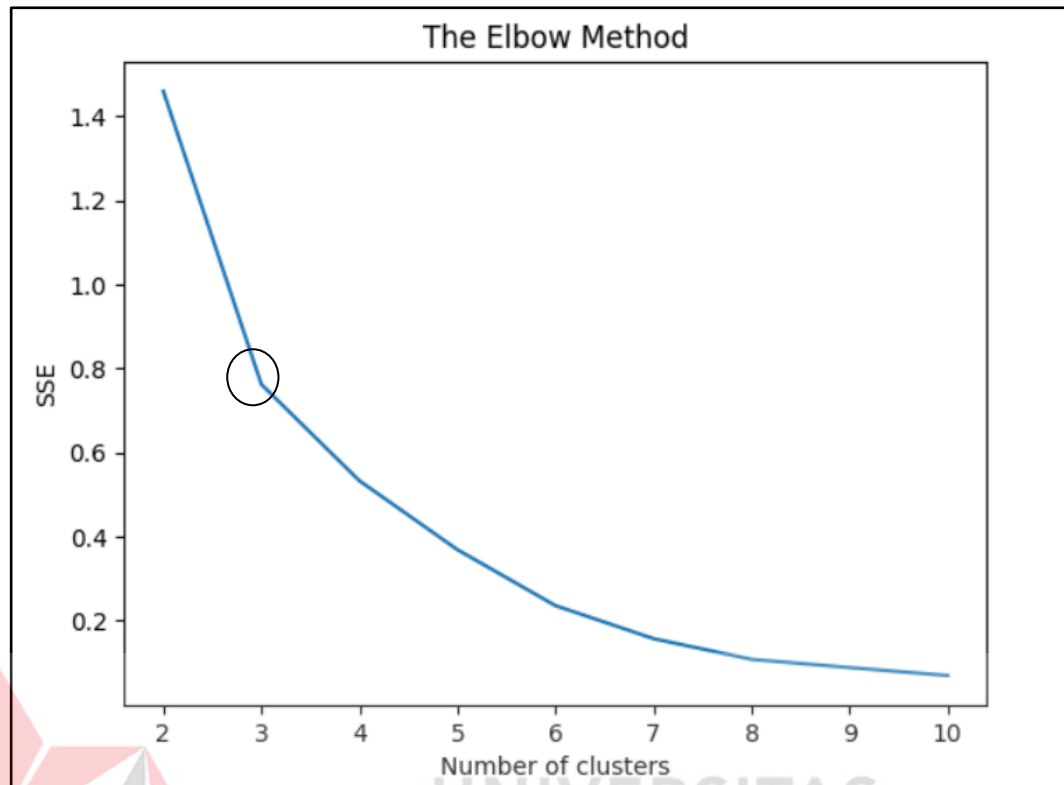
4.3.3 Uji Coba 3 dengan 25 Data Tempat Wisata

Dari hasil proses perhitungan *Sum of Square Error* terhadap 25 data tempat wisata maka hasil yang mengalami penurunan yang paling besar adalah pada $K=3$. Hasil dari perhitungan *Sum of Square Error* 25 data dapat dilihat pada Tabel 4.10.

Tabel 4.9 Hasil Perhitungan *Sum of Square Error* 25 Data

Cluster	Hasil <i>Sum Square of Error</i>	Selisih
2	1,4593	1.4593
3	0,7622	0,6971
4	0,5329	0,2293
5	0,3690	0,1639
6	0,2359	0,1331
7	0,1573	0,0786
8	0,1081	0,0492
9	0,0886	0,0195
10	0,0697	0,0189

Hasil dari grafik *Sum of Square Error* 25 data dapat dilihat pada Gambar 4.4.



Gambar 4.4 Grafik *Sum of Square Error* 25 Data

Dari hasil proses perhitungan *Sum of Square Error* terhadap 25 data tempat wisata maka hasil grafik yang mengalami penurunan yang paling besar adalah pada $K=3$.

4.3.4 Perhitungan *K-Means Clustering* $K=3$

Dari ketiga uji coba yang telah dilakukan berdasarkan studi kasus data tempat wisata tahun 2021 yang diambil pada Dinas Kebudayaan dan Pariwisata Kabupaten Bojonegoro, hasil *Sum of Square Error* yang mengalami penurunan paling besar adalah pada $K=3$ dengan jumlah data yang berbeda. Sehingga untuk kasus ini jumlah *cluster* yang ideal adalah $K=3$ dan dijadikan sebagai *cluster* untuk menentukan pengelompokan dari 25 tempat wisata. Berikut ini adalah tahapan proses perhitungan *k-means clustering* $K=3$

1. Proses mengimport *library*

```
import numpy as np
import pandas as pd
from sklearn.cluster import KMeans
```

NumPy digunakan untuk proses manipulasi data dan mengoperasikan *array*, *pandas* digunakan untuk proses membaca data set yang akan diolah, sedangkan *sklearn.cluster* digunakan untuk proses *k-means clustering*.

2. Proses *insert dataset* berformat csv yang akan diolah

```
dataset = pd.read_csv("wisata_fix.csv")
data = dataset.iloc[:, 1:5]
data.head(25)
```

Tabel 4.10 Hasil *Insert Dataset*

Nama Wisata	Persentase Jumlah Kunjungan	Persentase Luas Wilayah
Dander_Waterpark	0.639	0.83
Kayangan_Api	0.566	0.25
Waduk_Pacal	0.793	0.80
Taman_Pinggir_Gawan	0.423	0.19
Negeri_Atas_Angin	0.838	0.78
...

3. Proses menentukan titik *centroid*

```
centroids = kmeans.cluster_centers_
print("Centroid:")
for centroid in centroids:
    print(centroid)
```

```
Centroid:
[0.5294 0.336]
[0.42488889 1.00]
[0.782 0.81181818]
```

4. Proses inialisasi $k=3$, membuat objek *k-means* dengan jumlah kelompok yang diinginkan, serta melakukan proses *clustering* data

```
k = 3
kmeans = KMeans(n_clusters=k)
y_predicted = kmeans.fit(df)
```

5. Proses mendapatkan label kelompok dari setiap data

```
labels = kmeans.labels_
df['Cluster']= labels
df
```

Tabel 4.11 Hasil Pelabelan Setelah Proses *Clustering*

Nama Wisata	Persentase Jumlah Kunjungan	Persentase Luas Wilayah	Cluster
Dander_Waterpark	0.639	0.83	2
Kayangan_Api	0.566	0.25	0
Waduk_Pacal	0.793	0.80	2
Taman_Pinggir_Gawan	0.423	0.19	0

Nama Wisata	Persentase Jumlah Kunjungan	Persentase Luas Wilayah	Cluster
Negeri_Atas_Angin	0.838	0.78	2
...

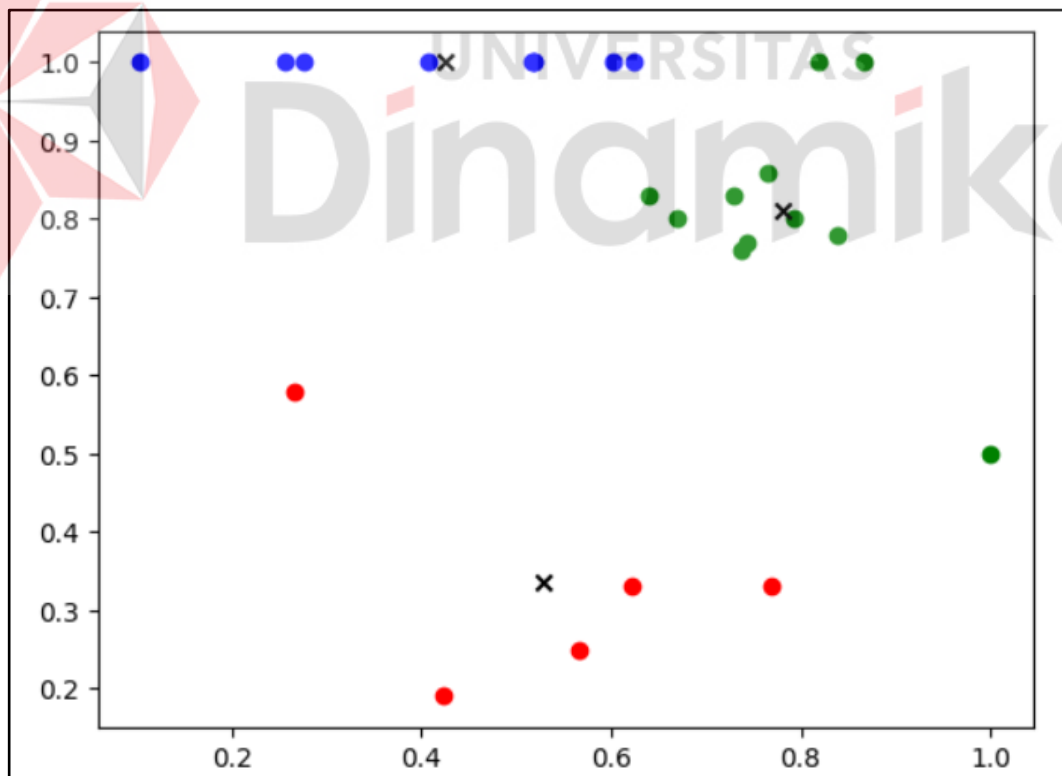
6. Proses visualisi *scatter* menggunakan *library matplotlib.pyplot* untuk melihat persebaran data

```
import matplotlib.pyplot as plt

dat0 = df[(df["Cluster"] == 0)]
dat1 = df[(df["Cluster"] == 1)]
dat2 = df[(df["Cluster"] == 2)]

plt.scatter(dat0["Persentase Jumlah Kunjungan"] ,
            dat0["Persentase Luas Wilayah"] , color = 'red')
plt.scatter(dat1["Persentase Jumlah Kunjungan"] ,
            dat1["Persentase Luas Wilayah"] , color = 'blue')
plt.scatter(dat2["Persentase Jumlah Kunjungan"] ,
            dat2["Persentase Luas Wilayah"] , color = 'green')

plt.scatter(centroids[:,0], centroids[:,1], marker="x",
            color='black')
plt.show()
```



Gambar 4.5 Hasil *Scatter*

7. Proses menghitung banyaknya jumlah anggota dari setiap *cluster*

```
dataset["Cluster"]=labels
df["Cluster"].value_counts()
```

Tabel 4.12 Jumlah Anggota Dari Setiap *Cluster*

Cluster	Jumlah
0	5
1	9
2	11

Setelah melalui proses tahapan perhitungan dengan *k-means clustering*, berikut ini adalah hasil rekap anggota *cluster* 0, *cluster* 1, dan *cluster* 2 yang dapat dilihat pada tabel 4.15

Tabel 4.13 Hasil Rekap Anggota *Cluster*

<i>Cluster</i>	<i>Anggota Cluster</i>	Persentase Jumlah Kunjungan (%)	Persentase Luas Wilayah (%)
0	Agrowisata Jambu Padang	27.7%	100%
	Wisata Edukasi Gerabah	40.7%	100%
	Wali Kidangan	51.8%	100%
	Sendang Tirta Arum	51.8%	100%
	Rumah Tua Padangan	25.6%	100%
	Puthuk Kreweng	62.5%	100%
	Masyarakat Samin	60.2%	100%
	Agrowisata Salak Wedi	10.4%	100%
	Petilasan Angling Dharma	51.7%	100%
1	Kayangan Api	56.6%	25%
	Taman Pinggir Gawan (TPG)	42.3%	19%
	Agrowisata Belimbing	62.2%	33%
	Waduk Bendo	76.9%	33%
	Penangkaran Rusa Malo	26.7%	58%
2	Dander Waterpark	63.9%	83%
	Waduk Pacal	79.3%	80%
	Negeri Atas Angin	83.8%	78%
	Teksas Wonocolo	67%	80%
	Rumah Singgah	81.8%	100%
	Growgoland	100%	50%
	Mumbul Kalianyar	86.7%	100%
	Banyu Kuning	74.4%	77%
	Outbond Toyoaji	73.8%	76%
	Wisata Edukasi Pejambon	76.6%	86%
	Sumber Maha Dewi	72.9%	83%

Berdasarkan hasil perhitungan *k-means clustering* dengan $K=3$ maka didapatkan karakteristik dari anggota setiap *cluster* sebagai berikut

Tabel 4.14 Karakteristik Anggota Setiap *Cluster*

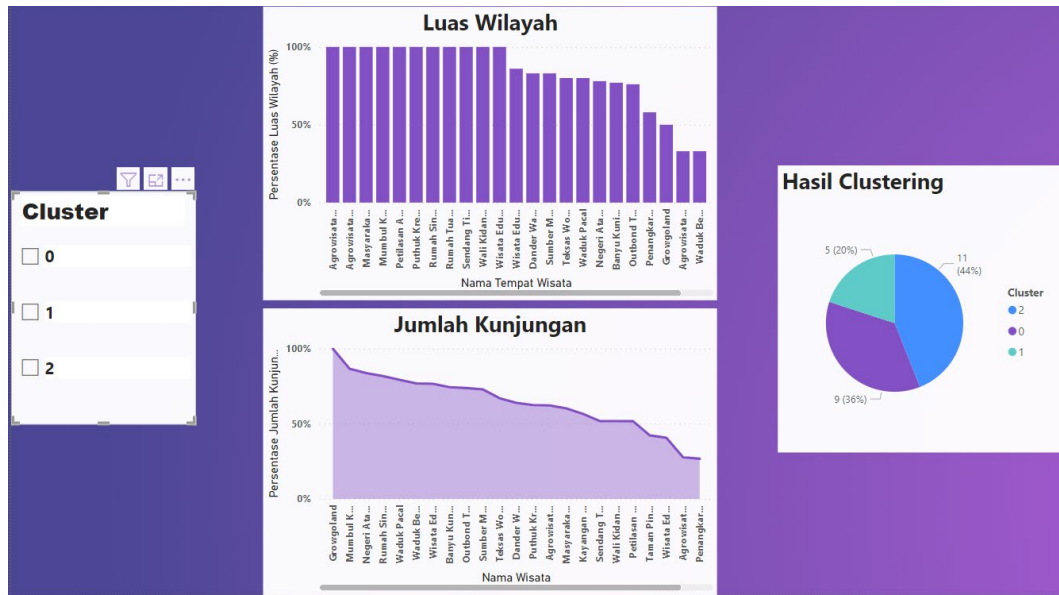
<i>Cluster</i>	Karakteristik
0 (Merah)	Persentase jumlah kunjungan cenderung rendah daripada <i>cluster</i> 1 dan <i>cluster</i> 2 sedangkan persentase luas wilayah sudah mencapai batas maksimal

1 (Biru)	Persentase jumlah kunjungan cenderung lebih tinggi daripada <i>cluster</i> 0 sedangkan persentase luas wilayah masih kecil
2 (Hijau)	Persentase jumlah kunjungan paling tinggi daripada <i>cluster</i> 1 dan <i>cluster</i> 2 begitu juga dengan persentase luas wilayah yang tinggi juga bahkan hampir maksimal

Berdasarkan karakteristik anggota setiap *cluster* tersebut, maka *cluster* 0 tergolong kelompok *low potential potential* di mana Dinas Kebudayaan dan Pariwisata Kabupaten Bojonegoro dapat melakukan promosi wisata dengan cakupan yang lebih luas agar dapat menarik kunjungan wisatawan, *cluster* 1 tergolong kelompok *high potential* di mana Dinas Kebudayaan dan Pariwisata Kabupaten Bojonegoro dapat memaksimalkan pemanfaatan luas wilayah tempat wisata sehingga dapat menarik jumlah kunjungan wisata lebih banyak karena meskipun untuk sekarang belum maksimal tetapi jumlah kunjungan termasuk ke dalam jumlah yang cukup tinggi, sedangkan *cluster* 2 tergolong kelompok *high priority* di mana Dinas Kebudayaan dan Pariwisata Kabupaten Bojonegoro diupayakan dapat mempertahankan atau justru lebih meningkatkan pengelolaan daya tarik wisata untuk mendapatkan jumlah kunjungan yang lebih tinggi karena masih ada beberapa tempat wisata yang masih dapat dikembangkan lagi dari sisi luas wilayah.

4.3.5 Pembuatan Visualisasi Data dengan *Power BI*

Fungsi dari visualisasi yaitu untuk memberikan informasi yang berdasarkan pada pengolahan data dan juga hasil *clustering* dalam bidang pariwisata sehingga pengguna lebih mudah dalam membaca hasilnya. Dashboard dapat dilihat pada Gambar 4.6.



Gambar 4.6 *Dashboard* Hasil Pengolahan Data dengan $K=3$

Pada gambar 4.6 terdapat *slicer* di sebelah kiri yang dapat dipilih untuk menampilkan data baik dari *cluster 0*, *cluster 1*, maupun *cluster 2*. Sedangkan *pie chart* menampilkan persentase dari hasil *clustering*. Kemudian untuk *bar chart* dan *line chart* menampilkan hasil data yang sudah diolah. Hasil tersebut divisualkan berdasarkan perhitungan *k-means clustering* dengan menampilkan data sesuai *cluster* serta karakteristik dari anggotanya.

BAB V PENUTUP

5.1 Kesimpulan

Berdasarkan hasil implementasi data mining menggunakan *k-means clustering* dan evaluasi dengan tahapan testing yang sudah dilakukan didapatkan hasil kesimpulan sebagai berikut:

1. Berdasarkan hasil uji coba baik 10 data, 15 data, maupun 25 data didapatkan hasil *cluster* yang optimal adalah $K=3$ atau pengelompokan data menjadi 3 *cluster*.
2. Bidang Pariwisata dapat mengetahui pengelompokan tempat wisata berdasarkan proses olah data menggunakan *k-means clustering* dengan nama *cluster high priority, high potential, dan low potential*.
3. Bidang Pariwisata dapat melihat hasil olah data pada tahun 2021 dengan jumlah *cluster* $K=3$ yang telah dibuat melalui visualisasi menggunakan *Power BI*.

5.2 Saran

Hasil implementasi data mining menggunakan *k-means clustering* masih memiliki kekurangan, sehingga terdapat saran yang dapat diimplementasikan untuk penelitian selanjutnya yaitu:

1. Data yang digunakan bisa lebih banyak dengan melakukan *clustering* terhadap tempat-tempat wisata yang berada di sekitar Kabupaten Bojonegoro.
2. *Dashboard* yang dibuat dapat lebih dilengkapi dalam hal tampilan data agar informasi yang disampaikan lebih maksimal

DAFTAR PUSTAKA

- Ayu, D., Dewi, I. C., & Pramita, K. (2019). Analisis Perbandingan Metode Elbow dan Silhouette pada Algoritma Clustering K-Medoids dalam Pengelompokan Produksi Kerajinan Bali. *Jurnal Matrix*, 9(3), 102.
- Edwin Bororing, E. (2022). Implementasi Dashboard Microsoft Power BI Untuk Visualisasi Data Covid 19 Indonesia. *Jurnal Informasi Interaktif*, 7.
- Gowthami, K., & Pavan Kumar, M. R. (2017). Study on Business Intelligence Tools for Enterprise Dashboard Development. *International Research Journal of Engineering and Technology*. www.irjet.net
- H. Witten, I., Frank, E., A. Hall, M., & J. Pal, C. (2011). *Data Mining: Practical Machine Learning Tools and Techniques*.
- Han, J., Kamber, M., & Pei, J. (2011). *Data Mining. Concepts and Techniques, 3rd Edition (The Morgan Kaufmann Series in Data Management Systems)*.
- Hanafi Ahmad, A. (2022). Pengaruh Jumlah Kunjungan Wisatawan, Objek Wisata, Dan Retribusi Pariwisata Terhadap Pendapatan Asli Daerah. *Jurnal Sosial Ekonomi Bisnis*, 2(1), 50–61. <https://doi.org/10.55587/jseb.v2i1.34>
- Harahap, B. (2017). *Penerapan Algoritma K-Means Untuk Menentukan Bahan Bangunan Laris (Studi Kasus Pada UD. Toko Bangunan YD Indarung)*.
- Imam Iskandar, M., & Ariffudin Islam, M. (2021). Destination Branding: Pendekatan Segmentation, Targetting, dan Positioning Dalam Perancangan Media Wisata Bahari Tlocor Sidoarjo. *Jurnal Barik*, 1(3), 164–180. <https://ejournal.unesa.ac.id/index.php/JDKV/>
- Katja, S., & Maufrais, C. (2010). *Introduction to Programming using Python*. <http://www.pearsonhighered.com/liang>
- Lusianah, N., Irma Purnamasari, A., & Nurhakim, B. (2023). Implementasi Algoritma K-Means Dalam Pengelompokan Jumlah Wisatawan Akomodasi Di Jawa Barat. *Jurnal Ekonomi, Bisnis Dan Manajemen*.
- Ncr, (, & Clinton, J. (1999). *CRISP-DM 1.0 Step-by-step data mining guide*. DaimlerChrysler.

- Pamungkas, F., Nugroho, D., & Utami, Y. R. W. (2020). Rekomendasi Wisata Umbul dengan K-Means Clustering. *Jurnal Teknologi Informasi Dan Komunikasi (TIKomSiN)*, 8(2). <https://doi.org/10.30646/tikomsin.v8i2.478>
- Putu, N., Merliana, E., & Santoso, A. J. (2017). *Analisa Penentuan Jumlah Cluster Terbaik Pada Metode K-Means Clustering*.
- Rohmah, A., Sembiring, F., & Erfina, A. (2021). *Implementasi Algoritma K-Means Clustering Analysis Untuk Menentukan Hambatan Pembelajaran Daring (Studi Kasus: SMK Yaspim Gegerbitung)*. <https://www.alfasoleh.com/2019/11/k-means-clustering-contoh>
- Romzi, M., & Kurniawan, B. (2020). *Pembelajaran Pemrograman Python Dengan Pendekatan Logika Algoritma* (Issue 2).
- Savitri, N., Pranata, R., Nadzario, A., Clara, M., Rahajeng, S., Politeknik, J. S., & Stis, S. (2021). Pengelempokan Kunjungan Wisata Kabupaten Kulon Progo Tahun 2019 Menggunakan K-Means Clustering. *Just IT: Jurnal Sistem Informasi, Teknologi Informasi Dan Komputer*, 12(1), 38–45. <https://jurnal.umj.ac.id/index.php/just-it/index>



UNIVERSITAS
Dinamika