

BAB 3

METODOLOGI PENELITIAN

3.1 Subjek dan Objek Penelitian

Bagian ini membahas terkait objek dan subjek penelitian yang akan digunakan. Adapun objek dan subjek penelitian sebagai berikut:

3.1.1 Subjek Penelitian

Subjek penelitian merupakan sumber-sumber informasi mengenai data penelitian diperoleh. Subjek penelitian adalah individu atau kelompok yang menjadi fokus utama dalam sebuah penelitian. Subjek penelitian ini dipilih berdasarkan tujuan penelitian. Subjek penelitian yang akan digunakan adalah Dinas Pemuda, Olahraga, Kebudayaan dan Pariwisata Kabupaten Tegal.

3.1.2 Objek Penelitian

Objek penelitian merujuk pada sesuatu yang menjadi titik perhatian di dalam penelitian menjadi hal yang diteliti. Objek penelitian juga harus relevan dengan masalah yang ingin dipecahkan atau pertanyaan penelitian yang ingin dijawab. Penelitian ini menggunakan destinasi wisata yang terdapat di wilayah Kabupaten Tegal sebagai objek penelitiannya.

3.2 Alat dan Bahan Penelitian

Penelitian ini dilakukan menggunakan beberapa alat dan bahan sebagai peninjau keberhasilan dalam penelitian ini.

3.2.1 Alat Penelitian

Perangkat keras dan perangkat lunak digunakan dalam penelitian ini untuk mencapai hasil akhir. Berikut adalah detail dari masing-masing perangkat yang digunakan dalam penelitian ini:

3.2.1.1 Perangkat Keras (*Hardware*)

- a. Laptop LENOVO Ideapad 330
- b. *Mouse* standar

3.2.1.2 Perangkat Lunak (*Software*)

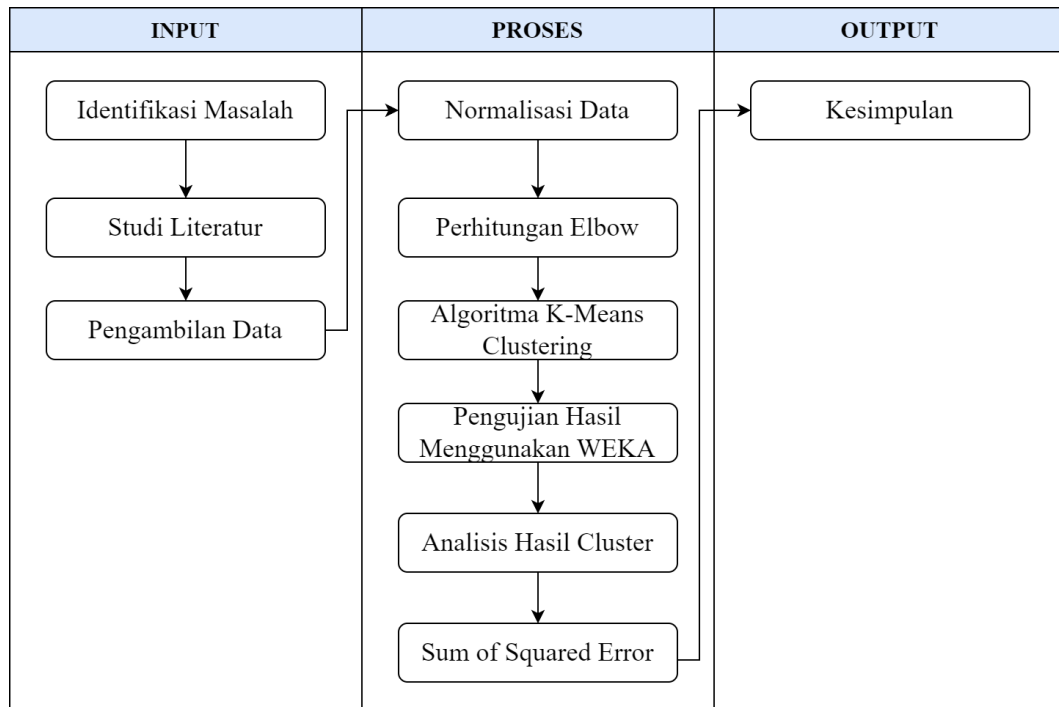
- a. Sistem Operasi Windows 10
- b. *Text editor* Notepad
- c. *Google chrome dan edge*
- d. *Microsoft Office Word 2019*
- e. *Microsoft Office Excel 2019*
- f. *Draw.io*
- g. *Mendeley*
- h. *Software WEKA*

3.2.2 Bahan Penelitian

Bahan yang digunakan dalam penelitian ini berupa data informasi data tentang objek wisata di Kabupaten Tegal. Penelitian ini menggunakan beberapa data, antara lain data destinasi wisata, data kunjungan wisata, data harga tiket masuk wisata dan jumlah fasilitas yang ada di tempat destinasi wisata.

3.3 Diagram Alir Penelitian

Diagram alir penelitian adalah suatu visualisasi yang menunjukkan urutan langkah-langkah atau proses yang dilakukan dalam penelitian. Diagram alir penelitian digunakan untuk mempermudah pemahaman alur penelitian yang dilakukan. Diagram alir penelitian dapat membantu peneliti dalam merencanakan dan mengatur penelitian secara sistematis, sehingga dapat meminimalkan kesalahan dan memastikan kesesuaian antara tahapan penelitian yang dilakukan. Pada proses penelitian ini, identifikasi masalah akan dilakukan terlebih dahulu, kemudian dilakukan studi literatur yang relevan dengan masalah, mengumpulkan data dilanjutkan dengan proses normalisasi data menggunakan metode *Z-Score Normalization*, kemudian melanjutkan proses algoritma *k-means clustering*, pengujian data menggunakan aplikasi *WEKA*, menganalisis hasil *cluster*, kemudian melakukan penghitungan *Sum of Squared Error (SSE)* dan menyimpulkan hasil berupa rekomendasi. Proses kegiatan mengenai urutan yang akan dilakukan dapat dilihat pada Gambar 3.1.



Gambar 3. 1 Diagram Alir Penelitian

3.3.1 Identifikasi Masalah

Pada tahap ini, dilakukan penentuan dan formulasi masalah pada penelitian dengan tujuan untuk mengidentifikasi masalah yang sedang dihadapi di Kabupaten Tegal melalui wawancara, maka diketahui masalah yang ada di Kabupaten Tegal yaitu belum adanya informasi yang jelas mengenai pengelompokan destinasi wisata untuk rekomendasi wisatawan.

3.3.2 Studi Literatur

Pada langkah ini, dilakukan *review* pustaka, yaitu mengkaji referensi seperti jurnal, artikel, buku, dan sumber lainnya. *Review* pustaka bertujuan untuk memperoleh dasar teori tentang penelitian yang akan dilakukan, seperti teknik pengolahan data *mining* dan proses pengolahan data *Clustering* menggunakan algoritma *K-Means*. Tujuan dari *review* pustaka adalah untuk memahami metode yang akan diteliti secara tepat.

3.3.3 Pengumpulan Data

Tahap ini melibatkan pengumpulan data yang diperlukan untuk penelitian yang akan dilaksanakan, seperti data destinasi wisata yang ada di Kabupaten Tegal, harga tiket masuk, data banyaknya kunjungan wisatawan, dan banyaknya fasilitas pada destinasi wisata. Untuk memperoleh data dalam penelitian ini, digunakan teknik pengumpulan data secara sekunder, wawancara dan penelitian kepustakaan. Wawancara dilakukan dengan pihak Dinas Pemuda, Olahraga, Kebudayaan, dan Pariwisata Kabupaten Tegal untuk menggali fakta mengenai parameter yang akan digunakan dalam penelitian ini. Penelitian kepustakaan yang dilakukan seperti *Library research* untuk mencari data-data dari buku maupun jurnal penelitian yang relevan dengan topik penelitian yang akan dilaksanakan.

3.3.4 Proses Normalisasi Data

Tahapan ini melakukan normalisasi data sebelum melakukan proses *k-means clustering*. Data yang akan dinormalisasi pada tahapan ini merupakan data jumlah fasilitas wisata, harga tiket wisata, dan jumlah kunjungan wisata. Untuk melakukan normalisasi pada data yang telah terkumpul, digunakan teknik *Z-Score Normalization*. Setelah menentukan teknik yang akan digunakan, kemudian menghitung nilai minimum dan maksimum dari data yang akan digunakan. Setelah menghitung nilai minimum dan maksimum, kemudian menghitung nilai normalisasi untuk setiap data dengan persamaan normalisasi dengan jenis normalisasi yang dipilih. Berikut ini merupakan contoh perhitungan normalisasi data menggunakan *z-score normalization*:

Misalkan kita memiliki data penghasilan bulanan sejumlah 5 orang karyawan dalam juta rupiah sebagai berikut:

5, 7, 9, 10, 14

- a. Langkah pertama menghitung rata-rata dari data tersebut

$$\bar{X} = (5 + 7 + 9 + 10 + 14)/5$$

$$\bar{X} = 45/5$$

$$\bar{X} = 9$$

b. Menghitung nilai standar deviasi (σ)

$$std_{dev} = \sqrt{\left(\frac{((5-9)^2 + (7-9)^2 + (9-9)^2 + (10-9)^2 + (14-9)^2)}{5}\right)}$$

$$std_{dev} = \sqrt{\left(\frac{((4)^2 + (-2)^2 + (0)^2 + (1)^2 + (5)^2)}{5}\right)}$$

$$std_{dev} = \sqrt{\left(\frac{(16 + 4 + 0 + 1 + 25)}{5}\right)}$$

$$std_{dev} = \sqrt{\left(\frac{43}{5}\right)}$$

$$std_{dev} = \sqrt{8.6}$$

$$std_{dev} = 2.9325$$

c. Melakukan perhitungan normalisasi data pada pertama (5)

$$X_{baru} = \frac{X_{lama} - \bar{X}}{\sigma}$$

$$X_{baru} = \frac{5 - 9}{2.9325}$$

$$X_{baru} = \frac{-4}{2.9325}$$

$$X_{baru} = -1.3640$$

Hasil perhitungan normalisasi data yang diperoleh dari contoh penghasilan bulanan karyawan pada data pertama sebesar **-1.3640**. Perhitungan normalisasi data juga dapat dilakukan menggunakan *Microsoft Excel* untuk mempermudah dan mempercepat perhitungan normalisasi data.

3.3.5 Perhitungan *Elbow*

Tahapan ini dilakukan perhitungan *Elbow* untuk menentukan jumlah *cluster* terbaik untuk pengelompokan data. Metode yang digunakan untuk menganalisis *cluster* yaitu metode *Elbow* untuk menentukan jumlah optimal dari *cluster*. Metode *Elbow* digunakan dengan memperhatikan nilai perbandingan antara jumlah *cluster* yang dibentuk dengan nilai *SSE* yang dihasilkan. Nilai *SSE* akan semakin mengecil

seiring dengan bertambahnya jumlah *cluster* k yang dibentuk, hingga bentuk siku pada grafik terbentuk dan menunjukkan jumlah *cluster* yang optimal.

3.3.6 Metode Algoritma *K-Means Clustering*

Proses algoritma *K-Means Clustering* dilakukan dengan beberapa tahapan, berikut penjelasan mengenai tahapan-tahapan algoritma *K-Means Clustering*:

- 1) *Pre-processing (normalisasi)*, pada tahap ini dilakukan penyaringan data yang akan digunakan untuk mengelompokkan data. Tahapan ini menggunakan persamaan *Z-Score Normalization* untuk menormalisasikan data.
- 2) Tentukan jumlah *cluster*, Pada tahap ini, dilakukan penentuan jumlah klaster k yang akan dibuat dari *dataset* yang ada. *Cluster* yang digunakan dalam penelitian ini terdapat 3 *cluster*.
- 3) Pusat *cluster (centroid)*, pada tahap ini dilakukan penentuan secara acak data yang menjadi pusat awal *cluster*. Pusat *cluster* yang akan digunakan untuk perhitungan jarak objek menggunakan data ke-39 untuk *cluster* 0 (C0), data ke-43 untuk *cluster* 1 (C1), dan data ke-53 untuk *cluster* 2 (C2).
- 4) Hitung jarak objek ke pusat *cluster*, Pada tahapan ini dilakukan perhitungan jarak antara objek dengan pusat *cluster* untuk menentukan kelompok data. Perhitungan jarak menggunakan metode *Euclidean Distance*. Berikut ini merupakan contoh perhitungan jarak objek ke pusat *cluster* menggunakan persamaan 2.3:

Misalnya terdapat data

$x = 1.643575$, $y = 0.440440$, $z = 7.183614$ dengan

Nilai pusat *cluster* 0 = 0.425086, 1.068011, -0.151902

Nilai pusat *cluster* 1 = 0.234698, -0.814701, -0,070991

Nilai pusat *cluster* 2 = 1.795886, 1.068011, -00.064682

Perhitungan jarak data ke pusat *cluster* menggunakan data contoh, dapat dilihat sebagai berikut:

Jarak data ke-1 ke *cluster* 0

$$\begin{aligned}
 &= \sqrt{((1.643575 - 0.425086)^2 + (0.440440 - 1.068011)^2 + (7.183614 - (-0.151902))^2)} \\
 &= \sqrt{1.484715 + 0.393845 + 53.809794}
 \end{aligned}$$

$$= \sqrt{55.688354}$$

$$= 7.462463$$

Jarak data ke-1 ke *cluster* 1

$$= \sqrt{((1.643575 - 0.234698)^2 + (0.440440 - (-0.814701))^2 + (7.183614 - (-0.070991))^2)}$$

$$= \sqrt{1.984934 + 1.575379 + 52.629293}$$

$$= \sqrt{56.189606}$$

$$= 7.495973$$

Jarak data ke-1 ke *cluster* 2

$$= \sqrt{((1.643575 - 1.795886)^2 + (0.440440 - 1.068011)^2 + (7.183614 - (-0.064682))^2)}$$

$$= \sqrt{0.023197 + 0.393845 + 52.537795}$$

$$= \sqrt{52.954837}$$

$$= 7.277007$$

- 5) Kelompokkan objek berdasarkan jarak terdekat, pada tahap ini dilakukan pengelompokkan data berdasarkan hasil perhitungan jarak objek ke pusat *cluster*.
- 6) Melakukan analisis kelompok data, pada tahap ini melakukan analisis terhadap kelompok data, apabila kelompok data sudah sama (*convergen*) maka proses *clustering* selesai. Apabila kelompok data belum sama maka melakukan proses penentuan pusat *cluster* kembali.

3.3.7 Pengujian Menggunakan WEKA

Tahapan ini dilakukan evaluasi terhadap hasil pengelompokan data yang telah dilakukan dengan menggunakan algoritma *k-means clustering* dan divalidasi menggunakan teknik *Sum of Squared Error (SSE)*. Pengujian menggunakan WEKA dilakukan untuk memeriksa seberapa baik algoritma dapat digunakan. Pengujian menggunakan WEKA dapat membantu untuk mengevaluasi algoritma yang dipilih.

3.3.8 Analisis Hasil Cluster

Tahapan ini dilakukan analisis data yang sudah dikelompokkan atau di-*clusterkan* untuk mengidentifikasi pola atau hubungan diantara kelompok-kelompok tersebut. Hasil dari analisis ini dapat memberikan wawasan atau

pemahaman baru terkait data yang diuji serta membantu dalam pengambilan keputusan berdasarkan pola atau hubungan yang ditemukan.

3.3.9 Proses Perhitungan *Sum of Squared Error (SSE)*

Tahapan ini dilakukan perhitungan nilai *Sum of Squared Error (SSE)* untuk mengukur seberapa baik pengelompokan data yang telah dilakukan dengan menggunakan algoritma *k-means clustering*. Hasil perhitungan jarak data ke pusat *cluster* dapat digunakan untuk melakukan perhitungan *Sum of Squared Error (SSE)*. Perhitungan *Sum of Squared Error* dapat dilakukan menggunakan persamaan 2.4, berikut ini merupakan salah satu perhitungan *Sum of Squared Error (SSE)*:

Perhitungan *Sum of Squared Error (SSE)* pada *cluster 0 (C0)*

$$\begin{aligned} SSE_0 &= (1.293180)^2 + (1.108773)^2 + (1.259332)^2 + (0.350788)^2 + (0.822889)^2 + \\ &\quad (1.0704906)^2 + (0.516595)^2 + (1.506379)^2 + (1.023570)^2 + (0.486272)^2 + \\ &\quad (0.368016)^2 + (1.079386)^2 + (0.868593)^2 + (0.971864)^2 + (0.262553)^2 + \\ &\quad (1.492157)^2 \\ &= 15.548911 \end{aligned}$$

3.3.10 Kesimpulan

Pada tahap ini, dilakukan kesimpulan dari hasil analisis data setelah melalui beberapa tahap *data mining* menggunakan metode algoritma *K-Means Clustering* terkait pengelompokan destinasi wisata di Kabupaten Tegal.