

## **BAB II**

### **TINJAUAN PUSTAKA**

#### **2.1. Teori**

Peneliti berusaha menguraikan teori-teori yang mendasari studi ini dengan merinci konsep-konsep yang akan digunakan serta mengevaluasi teori-teori yang akan diterapkan

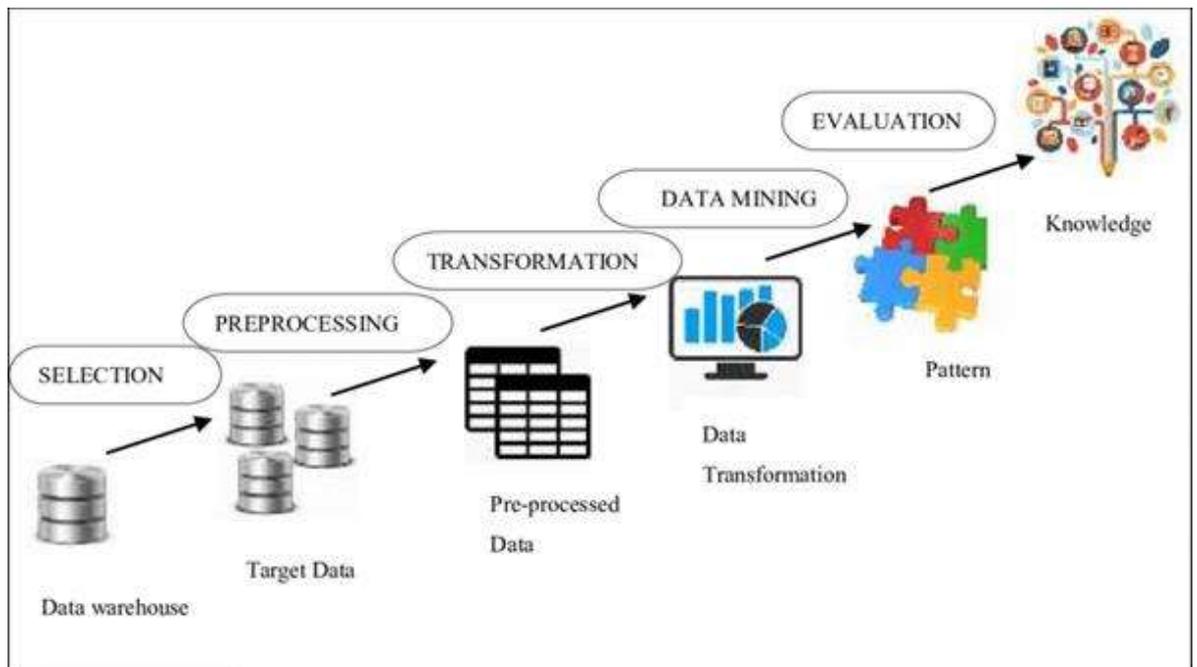
##### **2.1.1 Teori Umum**

Teori merupakan sebuah rumusan atau pernyataan yang berasal dari interpretasi seseorang terhadap fakta-fakta, atau penjelasan mengenai gejala-gejala yang terdapat dalam dunia fisik. Teori merupakan suatu abstraksi intelektual di mana pendekatan secara rasional digabungkan dengan pengalaman empiris. Teori yang baik memiliki kegunaan dalam hal teori tersebut dapat memberitahukan banyak hal kepada kita mengenai komunikasi dan perilaku manusia. Hal ini memungkinkan kita untuk mengetahui beberapa elemen dari komunikasi yang sebelumnya tidak jelas.

##### ***2.1.1.1 Knowledge Discovery in Database (KDD)***

Proses Ekstraksi ilmu dari Basis Data (KDD) merupakan salah satu metode yang diterapkan dalam melakukan data mining. Menurut Shirazi, Baziyad (2019) KDD adalah proses menggunakan teknik data mining untuk mengidentifikasi informasi berharga dan pola-pola yang tidak langsung terlihat dalam data dengan menggunakan algoritma khusus. Dunham (2018) merangkum proses KDD menjadi beberapa langkah, Termasuk di dalamnya

adalah pemilihan data, pra-pemrosesan data, transformasi data, eksplorasi data, serta interpretasi dan evaluasi hasil. Berikut adalah gambaran dan penjelasan mendetail mengenai proses KDD :



**Gambar 2.1** Tahapan *Knowledge Discovery in Database* (KDD)

*a. Selection*

Selection adalah proses di mana data diproses dan dipilih berdasarkan kriteria yang relevan untuk digunakan dalam tahap pemrosesan data mining. Sebelum memulai proses pencarian informasi dalam KDD, diperlukan seleksi data operasional untuk memilih data yang sesuai (Sundari et al., 2021).

*b. Processing*

*Processing* melibatkan penghapusan data duplikat dan identifikasi data yang unik. Proses ini mencakup pembuangan duplikasi data, pemeriksaan kesalahan data, dan perbaikan sebelum dilakukan proses data mining (Sundari et al., 2021).

### *c. Transformation*

*Transformation* merupakan proses untuk mengubah format atau karakteristik data tanpa mengubah esensi dari data yang telah ditetapkan dalam prosedur pengolahan data mining.

### *d. Data Mining*

*Data mining* adalah proses untuk menemukan informasi baru dan pola yang bermanfaat dari data. Proses ini melibatkan penggunaan berbagai teknik, metode, atau algoritma untuk mengekstraksi pola-pola yang potensial sesuai dengan tujuan KDD (Sundari et al., 2021).

### *e. Interpretation / Evaluation*

*Interpreation / Evaluation* melibatkan penafsiran corak informasi yang diperoleh agar dapat dipresentasikan dengan cara yang jelas dan dimengerti. Proses ini melibatkan evaluasi terhadap pola-pola yang telah diidentifikasi berdasarkan ukuran yang telah ditetapkan (Sundari et al., 2021).

### *f. Knowledge*

Pengetahuan adalah tahap terakhir dalam proses KDD, di mana data yang telah diproses divisualisasikan untuk membantu pengguna dalam memahami dan membuat keputusan berdasarkan analisis yang telah dilakukan. Gambaran ini menjelaskan proses-proses utama dalam KDD atau data mining, yang bertujuan untuk menggali potensi dari data yang tersedia, menganalisis pola, dan memvisualisasikan hasilnya agar dapat dipahami dan digunakan untuk membuat keputusan.

### 2.1.1.2 Data Mining

Data mining adalah proses untuk mengidentifikasi struktur dan keterkaitan yang tidak tampak dalam kumpulan data besar dengan tujuan untuk melakukan klasifikasi, estimasi, prediksi, identifikasi pola, pengelompokan, deskripsi, dan representasi visual. Secara umum, data mining dapat dibagi menjadi dua kategori utama:

- a. *Descriptive mining* yang bertujuan untuk mengidentifikasi fitur-fitur utama dari data yang disimpan dalam sebuah basis data.
- b. *Predictive mining* berfokus pada pembuatan model yang bertujuan untuk mengidentifikasi pola dalam data dan meramalkan nilai variabel di masa yang akan datang.

Berikut adalah langkah-langkah dalam prosedur data mining :

1. Pembersihan data (*data cleaning*) adalah langkah untuk menghapus gangguan serta mendeteksi dan memperbaiki data yang tidak konsisten atau tidak relevan
2. Integrasi data (*data integration*) adalah proses mengkombinasikan data dari berbagai sumber yang berbeda ke dalam satu basis data yang konsolidasi.
3. Seleksi data (*data selection*) dilakukan untuk menyaring dan memilah dan mengambil data yang signifikan serta diperlukan dari basis data untuk tahap analisis berikutnya
4. Transformasi data (*data transformation*) melibatkan perubahan atau penggabungan data ke dalam format yang lebih cocok untuk proses data mining.

5. Proses mining adalah tahapan utama di mana metode data mining diterapkan untuk mengeksplorasi dan menemukan pola atau wawasan berharga yang diperoleh dari informasi.
6. Evaluasi pola (*pattern evaluation*) diperuntukkan bagi mengevaluasi dan menemukan pola-pola signifikan yang terungkap selama proses data mining.
7. Presentasi pengetahuan (*knowledge presentation*) melibatkan visualisasi dan penyajian informasi mengenai metode dan temuan yang diperoleh dari proses data mining. ( Ridwan, 2018 )

### 2.1.1.3 K-Means

K-mean clustering adalah teknik pengelompokan yang tidak mengikuti struktur hierarkis, bertujuan untuk membagi data ke dalam beberapa kelompok. Pendekatan ini mengelompokkan data berdasarkan kesamaan karakteristik, sehingga data yang mirip dikumpulkan bersama, sementara data dengan perbedaan signifikan dipisahkan ke dalam kelompok yang berbeda (Prasetyo, 2018). Untuk menentukan jarak antara setiap data dan titik pusat cluster, dapat digunakan Rumus Euclidean yang dihitung dengan cara sebagai berikut:

$$D(i, j) = \sqrt{(x_{1i} - x_{1j})^2 + (x_{2i} - x_{2j})^2}$$

$D(I,j)$  = Jarak data ke i ke pusat cluster j

$X_{ki}$  = Data ke i pada atribut data ke k

$X_{kj}$  = Data ke j pada atribut data ke k

**Rumus 2. 1** *Rumus Euclidean*

#### 2.1.1.4 Clustering

*Clustering* adalah proses memisahkan atau memecah data ke dalam beberapa kelompok berdasarkan karakteristik tertentu. Pada *clustering*, label untuk setiap data belum diketahui pada awalnya, namun diharapkan bahwa setelah proses *clustering* selesai, kelompok data dapat diberi label sesuai keinginan. Algoritma ini dirancang untuk memudahkan pengelompokan data berdasarkan variabel tertentu tanpa memerlukan proses pelatihan (training) terlebih dahulu.

*K-means clustering* merupakan algoritma pembelajaran tanpa pengawasan yang berfokus pada penggunaan centroid untuk mengelompokkan data., di mana setiap kelompok diasosiasikan dengan sebuah centroid. Tujuan utama dari algoritma ini adalah untuk meminimalkan jarak antara titik data dan centroid dari kelompok yang sesuai. Algoritma *K-means* mengambil dataset yang tidak berlabel sebagai input, membaginya menjadi k cluster, dan mengulangi proses ini hingga konvergensi. Jumlah k (jumlah cluster) harus ditentukan sebelumnya dalam algoritma ini.

Algoritma *k-means clustering* melakukan dua tugas utama:

- 1) Mengidentifikasi nilai optimal untuk centroid k melalui iterasi.
- 2) Menetapkan setiap titik data ke centroid terdekat, membentuk cluster dari titik-titik data yang terdekat dengan centroid tertentu.

### 2.1.1.5 Titik Pusat / *Centroid*

Rumus centroid menurut Suhendra (2023) adalah metode yang digunakan dalam data mining untuk menentukan pusat data, atau centroid, dari suatu cluster. Pusat data ini digunakan sebagai representasi dari sekumpulan data dalam suatu cluster. Dalam menghitung pusat data, rumus centroid mempertimbangkan jarak antar data yang ada dalam cluster tersebut. Menentukan Centroid (Titik Pusat) setiap kelompok diambil dari nilai rata - rata (Means) semua nilai data pada setiap fiturnya. Jika  $M$  menyatakan jumlah data pada suatu kelompok,  $i$  menyatakan fitur ke- $i$  dalam sebuah kelompok, berikut rumus untuk menghitung centroid:

$$\mu_k = \frac{1}{N_k} \sum_{q=1}^{N_k} x_q$$

**Rumus 2. 2** Rumus *Centroid*

### 2.1.1.5 *Rapid miner*

Rapid miner adalah sebuah perangkat lunak ilmu data yang dikembangkan oleh perusahaan bernama sama. Perangkat lunak ini menyediakan lingkungan terintegrasi untuk persiapan data, pembelajaran mesin, penambangan teks, dan analisis prediktif (Dewi et al., 2019).

Rapid miner tersedia sebagai aplikasi *Open Source* yang dapat digunakan untuk melakukan datamining, text mining, dan analisis prediksi. Dengan 500 operator datamining, Rapidminer dapat berfungsi secara mandiri dalam menganalisis data (Ardiansyah & Walim, 2018).

## **2.1.2 Teori Khusus**

Penulis berusaha untuk menjelaskan berbagai teori yang relevan dengan situasi tertentu, serta menguraikan hubungan antara teori-teori tersebut dan masalah yang akan diteliti.

### **2.1.2.1 Travel Agent**

*Travel agent* adalah perusahaan yang menyediakan layanan perjalanan wisata beserta segala kelengkapannya (Faidah, 2022). Sebagai entitas bisnis, travel agent memiliki tujuan komersial untuk mencapai keuntungan. Untuk mencapai tujuan ini, ada beberapa aspek yang mendukung pengelolaan *travel agent*, di antaranya adalah strategi pemasaran. Strategi pemasaran sangat krusial dalam mengoperasikan sebuah perusahaan, terutama dalam industri *travel agent*, karena strategi ini mencakup rencana dan langkah-langkah untuk mencapai tujuan perusahaan di pasar serta untuk mengembangkan keunggulan kompetitif yang berkelanjutan di antara pesaing lainnya. Ada beberapa konsep strategi pemasaran yang dapat digunakan untuk mengelola dan mengevaluasi efektivitas pemasaran produk. Salah satu konsep yang relevan adalah bauran pemasaran (marketing mix), yang dapat diterapkan oleh *travel agent* untuk merencanakan serta mengevaluasi strategi pemasaran yang mereka jalankan. Hal ini sangat penting mengingat persaingan yang ketat di antara *travel agent* di kota Batam yang dikenal kaya akan destinasi wisata.

### **2.1.2.2 Promo**

Aini (2023) menjelaskan bahwa promo merupakan istilah yang umumnya dikenal, khususnya dalam kegiatan pemasaran. Secara sederhana, promo adalah

sebuah tawaran yang menguntungkan Definisi promo merujuk pada penawaran yang sering kali digunakan dalam kegiatan pemasaran untuk menyebarluaskan informasi mengenai produk kepada publik. Selain itu, promo ini merupakan strategi yang digunakan oleh pemasar untuk merangsang minat dan memengaruhi calon pembeli sehingga mereka tertarik pada produk yang ditawarkan. Tanpa adanya promosi, produk akan menghadapi kesulitan dalam menarik perhatian calon pembeli, mengingat peran penting promosi dalam meningkatkan penjualan. Promo juga dapat membantu membangun citra positif produk di mata calon konsumen serta membedakan produk dari pesaing. Harga promo ditetapkan lebih rendah dari harga normal untuk meningkatkan penjualan produk atau jasa tersebut.

Tujuan dari promosi adalah untuk menarik minat konsumen terhadap barang atau jasa yang disediakan oleh perusahaan. Produk yang dipromosikan sering kali adalah barang-barang baru yang belum dikenal luas di masyarakat. Produk promosi juga sering dibantu oleh petugas promosi yang terampil dalam meyakinkan konsumen, sehingga konsumen tertarik untuk membeli atau menggunakan produk tersebut. Jenis-jenis promo yang umumnya ditawarkan mencakup penawaran beli 1 gratis 1, kode diskon, penjualan kilat, paket bundling, dan penawaran cashback

### **2.1.2.3 Wisatawan**

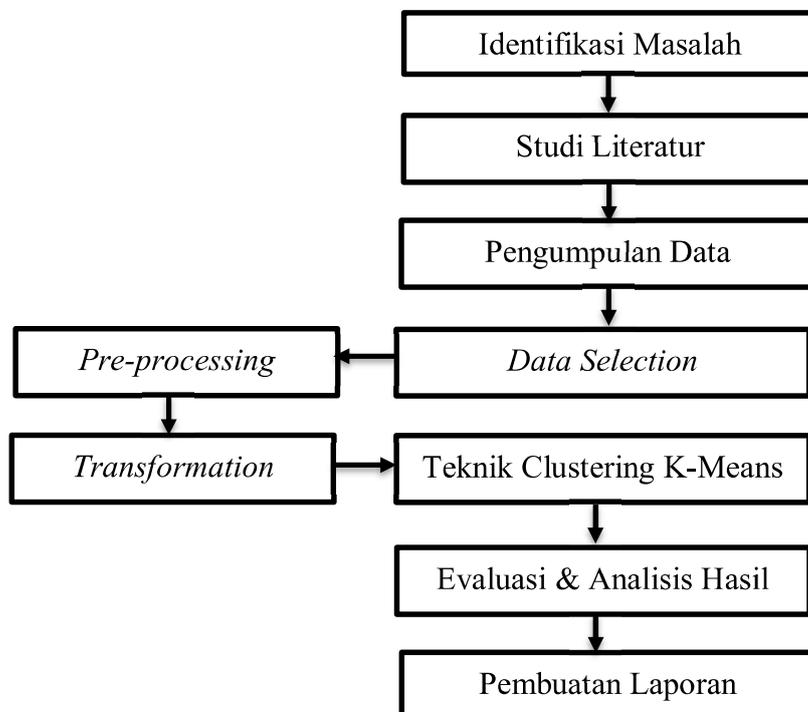
Wisatawan adalah individu atau kelompok yang melakukan perjalanan untuk tujuan beristirahat, berbisnis, berobat, melakukan kunjungan keagamaan, atau perjalanan studi. Dengan meninggalkan tempat tinggalnya untuk sementara waktu, seseorang dapat dianggap sebagai wisatawan. Selama melakukan

perjalanan tersebut, wisatawan memiliki maksud dan tujuan tertentu seperti beristirahat, berbisnis, dan lain-lain. Pentingnya batasan mengenai definisi wisatawan adalah untuk memahami jenis perjalanan dan ruang lingkup tempat wisata tersebut dilakukan. Wisatawan diklasifikasikan menjadi dua jenis:

1. Wisatawan Mancanegara adalah orang asing yang melakukan perjalanan wisata ke negara lain yang bukan tempat tinggalnya.
2. Wisatawan Lokal adalah warga negara yang melakukan perjalanan wisata di dalam wilayah negaranya sendiri tanpa melintasi batas negara.

## 2.2. Kerangka Pemikiran

Kerangka pemikiran berikut ini menyajikan rangkaian elemen yang menggambarkan urutan proses dalam penelitian ini. Berikut adalah kerangka berfikir yang digambarkan dalam metode penelitian ini :



**Gambar 2.2** Kerangka Pemikiran

Sumber: (Penulis, 2024)

### 2.3 Penelitian Terdahulu

Penulis menemukan beberapa studi terdahulu yang dapat mendukung penelitian terkait metode data mining dengan algoritma *clustering K-Means*.

1. Penelitian yang dilakukan oleh Pratama (2020) membahas “*Penerapan Metode Clustering Untuk Mengelompokkan Potensi Wisata Di Kabupaten Sumedang*”. Studi ini menunjukkan bahwa metode clustering dalam data mining adalah solusi efektif untuk mengatasi tantangan dalam mengelompokkan objek wisata. Dengan menggunakan algoritma *K-Means*, proses pengelompokan objek wisata dimulai dengan menentukan jumlah cluster, centroid masing-masing cluster, menghitung jarak antar data, dan mengelompokkan berdasarkan jarak minimum objek. Penerapan data mining dengan clustering dapat mempermudah identifikasi dan pengembangan rekomendasi objek wisata.

2. Lion (2023) melakukan penelitian tentang “*Penggunaan Algoritma K-Means Dalam Aplikasi Pemetaan Klaster Daerah Pariwisata*”. Penelitian ini bertujuan untuk mengelompokkan distrik-distrik potensial untuk pengembangan pariwisata menggunakan algoritma clustering. Algoritma *K-Means* menggunakan metode Elbow dan Silhouette untuk menentukan klaster yang optimal, yang kemudian direpresentasikan dalam peta wilayah pariwisata berbasis web. Studi ini menunjukkan hasil klasterisasi pada 7 distrik, dengan 3 distrik dalam klaster pertama dan 5 distrik dalam klaster kedua setelah normalisasi data kunjungan wisata.

3. Bambang (2021) memaparkan “*Penerapan Data Mining Untuk Mengelompokkan Kunjungan Wisatawan Di Kota Yogyakarta Dengan Metode*

*K-Means*". Penelitian ini menjelaskan bahwa k-Means ialah algoritma partitioning clustering yang efisien untuk memisahkan informasi ke dalam kluster terpisah. Melalui iterasi algoritma *K-Means*, studi ini berhasil menghasilkan 3 kluster dengan persentase data yang terbagi secara acak antara kluster 1 (50%), kluster 2 (23%), dan kluster 3 (27%).

4. Pattipeilohy (2022) melakukan penelitian tentang "*Implementasi K-Means Clustering dalam Analisis Data Mahasiswa Program Studi D4 Pariwisata untuk Merumuskan Strategi Pemasaran Yang Tepat*". Studi ini menggunakan *K-Means* untuk menganalisis dan mengelompokkan data mahasiswa dari berbagai provinsi, menunjukkan bahwa suplai mahasiswa terbanyak berasal dari Jawa Tengah, sementara beberapa provinsi lainnya memiliki kontribusi yang lebih rendah.

5. Rahmawati (2023) melakukan penelitian tentang "*Penerapan K-Means Clustering Pada Pariwisata Kabupaten Bojonegoro Untuk Mendukung Keputusan Strategi Pemasaran*". Penelitian ini menggunakan *K-Means* untuk mengelompokkan tempat wisata berdasarkan kondisi data yang ada, menghasilkan 3 kluster yang berbeda untuk digunakan dalam perumusan strategi pemasaran yang lebih efektif.

Dengan memanfaatkan temuan dari penelitian-penelitian tersebut, penulis dapat mengembangkan penelitian mereka sendiri yang berkaitan dengan penggunaan algoritma clustering *K-Means* dalam konteks data mining.