

BAB II

TINJAUAN PUSTAKA

2.1 Penelitian Terdahulu

Penelitian terkait merupakan salah satu acuan bagi peneliti dalam melakukan penelitian dan dapat menjadi referensi bagi penelitian yang dilakukan saat ini untuk mengefisienkan dan mengefektifkan pengembangan ilmu pengetahuan. Adapun penelitian terkait penelitian ini adalah sebagai berikut :

Penelitian yang dilakukan oleh Ahmad *et al.* [1] pada tahun 2025 dengan judul “Prediksi Harga Tiket Masuk dan Rating di Pulau Jawa Menggunakan Metode *DBSCAN*”. Penelitian ini berdasarkan penelitian yang dilakukan oleh peneliti di Kota Jawa dan metode yang digunakan adalah metode *K-Means* dan *DBSCAN*. Hasil penelitian menunjukkan bahwa algoritma *K-Means* memiliki kinerja yang lebih baik dengan skor *silhouette* sebesar 0,42, sedangkan skor *silhouette* algoritma *DBSCAN* sebesar 0,18. Penelitian ini bertujuan untuk menganalisis karakteristik destinasi wisata populer di Pulau Jawa menggunakan metode *K-Means* dan *DBSCAN*.

Rahmah Wulandari *et al.* [2] melakukan penelitian pada tahun 2022 dengan judul “*Analisis cluster berbasis informasi tekstual mengenai kebijakan pembukaan kawasan pariwisata selama pandemi*”. Penelitian ini mengabungkan metode *DBSCAN* dan *K-Means*. Hasil penelitian menunjukkan bahwa *DBSCAN* menghasilkan 2 *cluster* dan 9 *noise*, sedangkan *K-Means* menghasilkan 3 *cluster*.

Tujuan dari penelitian ini adalah untuk membandingkan kinerja algoritma *DBSCAN* dan algoritma *K-Means* dalam pengelompokan data.

Nurjannah *et al.* [3] melakukan penelitian dengan judul “Segmentasi Pengguna *E-Wallet* Menggunakan Metode *DBSCAN*” pada tahun 2023. Penelitian ini menggunakan metode *DBSCAN* dan quota sampling. Hasil penelitian menunjukkan bahwa *DBSCAN* melakukan klasterisasi berdasarkan jenis *e-wallet*, media pengisian ulang, dan tujuan penggunaan *e-wallet*, dengan tujuan untuk mengelompokkan pengguna dompet digital. Fitri Dwi Handayani *et al.* [4] melakukan penelitian dengan judul “Evaluasi Klasterisasi Pengguna Aplikasi Zenius pada Layanan *Google Play Store* Menggunakan Metode *DBSCAN*” pada tahun 2023. Hasil penelitian menunjukkan hasil rata-rata *silhouette coordinate correlation* (SC) yang berbeda-beda, dengan tujuan untuk memperoleh hasil evaluasi klasterisasi menggunakan metode *DBSCAN*.

Penelitian yang dilakukan oleh Putra Ratna Juwita Sari, [5] pada tahun 2020 dengan judul “*Clustering* rata-rata tingkat spending *mony* berdasarkan profil geografis dan prefekuensi konsumsi wisatawan mancanegara di kota Denpasar”, hasil dari penelitian ini terdapat variabilitas yang signifikan dalam tingkat pengeluaran wisatawan mancanegara di kota Denpasar yang dipengaruhi oleh profil geografis asal mereka. Penelitian ini perlu mengidentifikasi dan menganalisis faktor-faktor yang menyebabkan perbedaan tersebut.

2.2 Landasan Teori

2.2.1 Clustering (Pengelompokan data)

Menurut Han, [1] (2012), *clustering* merupakan proses pengelompokan sekumpulan objek sehingga objek-objek dalam kelompok yang sama (*cluster*) memiliki tingkat kemiripan yang tinggi satu sama lain. Dalam industri pariwisata, *clustering* digunakan untuk mengidentifikasi pola atau segmen perilaku wisatawan agar dapat menyesuaikan strategi layanan dan pemasaran secara efektif. Menurut Jain (2010), *clustering* sangat berguna dalam analisis data eksploratif, misalnya dapat digunakan untuk mengelompokkan konsumen berdasarkan perilaku atau preferensi belanjanya, atau dapat digunakan untuk mengelompokkan wisatawan berdasarkan struktur pengeluarannya.

2.2.2 Wisatawan

Wisatawan adalah suatu kegiatan yang melibatkan perjalanan orang dari satu tempat ke tempat lain dengan tujuan untuk menikmati dan mengalami keindahan alam, budaya, dan lain-lain. Pariwisata juga dapat diartikan sebagai suatu industri yang menyediakan jasa dan fasilitas untuk memenuhi kebutuhan wisatawan.

2.2.3 Komponen Pengeluaran

Menurut Kim [1] (2022), pengeluaran wisatawan merupakan seluruh biaya yang dikeluarkan selama perjalanan wisata, baik yang bersifat langsung maupun tidak langsung. Komponen utama pengeluaran wisatawan mencakup:

1. Akomodasi
2. Transportasi
3. Makanan dan Minuman

4. Tiket objek wisata
5. Belanja atau oleh-oleh
6. Lain-lain seperti jasa pemandu wisata

Menurut John [2] (2022) menyatakan bahwa pengeluaran wisatawan adalah bagian dari perjalanan yang dilakukan oleh individu ke destinasi yang berbeda dari tempat tinggalnya untuk tujuan rekreasi.

2.2.4 Algoritma DBSCAN

Algoritma *DBSCAN* menurut Han. [1] (2019). Mereka percaya bahwa *DBSCAN* merupakan sebuah algoritma *clustering* yang dapat mengelompokkan data berdasarkan kerapatan dan jarak antara data, serta dapat menangani data yang memiliki bentuk ukuran yang berbeda-beda.

Menurut Tan, Steinbach, dan Kumar [2] (2020), *DBSCAN* memiliki beberapa kelebihan dibandingkan metode *clustering* lain seperti *K-Means*:

1. Tidak memerlukan jumlah *cluster* sebagai parameter awal.
2. Dapat mendeteksi cluster yang bentuknya tidak beraturan.
3. Tahan terhadap *outlier* dan *noise*

Fitur ini sangat penting dalam *clustering* pariwisata karena data pengeluaran sering kali memiliki perubahan yang tidak normal dan nilai yang ekstrem (*outlier*), yang dapat mengganggu hasil analisis jika tidak ditangani dengan baik.

2.2.5 Penerapan DBSCAN

Menurut Rangkung [1] (2019), algoritma *clustering* berbasis *DBSCAN* cocok digunakan untuk analisis data pariwisata karena dapat mengelompokkan wisatawan berdasarkan perilakunya tanpa harus mendefinisikan secara eksplisit

jumlah kelompoknya. Hal ini penting ketika berhadapan dengan data eksploratif seperti pola konsumsi wisatawan yang kompleks dan beragam. Sementara itu, menurut Suharto [2] (2020), penggunaan algoritma *DBSCAN* untuk mengelompokkan wisatawan berdasarkan konsumsinya membantu pemerintah daerah atau pelaku industri pariwisata dalam merumuskan kebijakan harga, paket wisata, dan strategi pemasaran berdasarkan karakteristik masing-masing kelompok wisatawan.

2.2.5 Kekurangan dan kelebihan

Salah satu kelemahan utama *DBSCAN* adalah bergantung pada dua parameter penting, *eps* (jarak maksimum antara dua titik untuk dianggap bertetangga) dan *min_samples* (jumlah minimum titik yang dibutuhkan untuk membentuk kluster). Menentukan nilai parameter terbaik bisa sangat sulit, terutama ketika data memiliki perbedaan densitas yang signifikan antar kluster. Jika nilai *eps* terlalu kecil, banyak titik akan dianggap *noise*. Selain itu, *DBSCAN* tidak cocok untuk data dengan densitas yang bervariasi. Kekuatannya terletak pada kemampuannya menemukan kluster dengan bentuk yang tidak teratur atau kompleks. Selain itu, *DBSCAN* tidak mengharuskan pengguna untuk menentukan jumlah kluster di awal, karena jumlah kluster ditentukan secara optimis berdasarkan distribusi data dan parameter yang digunakan.

2.2.6 Data Mining

Istilah data mining sebenarnya mulai dikenal sejak tahun [1] 1990, ketika pekerjaan pemanfaatan data menjadi sesuatu yang dianggap penting dalam berbagai bidang, mulai dari bidang akademik, bisnis, hingga bidang medis. Munculnya data

mining didasarkan pada jumlah data yang tersimpan dalam basis data semakin besar. Perkembangan yang cepat dalam teknologi pengumpulan dan penyimpanan data telah memudahkan organisasi untuk mengumpulkan sejumlah data berukuran besar sehingga menghasilkan gunung data. Misalnya dalam sebuah universitas ada berapa data mahasiswa yang tersimpan dari tahun ke tahun. Kemudian data di sebuah supermarket, ada berapa transaksi pelanggan yang terjadi dalam sehari dan ada berapa juta data yang tersimpan dalam sebulan. Ekstraksi informasi yang berguna dari basis data tersebut menjadi pekerjaan yang cukup menantang. Seringkali alat dan teknik analisis data tradisional tidak dapat digunakan dalam mengekstrak informasi dari data berukuran 14 besar. Data mining adalah teknologi yang merupakan campuran teknik-teknik analisis data dengan algoritma untuk memproses data berukuran besar.

2.2.7 *Google Colab*



Gambar 2. 1 *Google Colab*

Google Colab, singkatan dari *Google Colaboratory*, adalah platform cloud yang disediakan oleh *Google* untuk menjalankan dan menulis kode *Python* melalui browser tanpa memerlukan konfigurasi tambahan. Ini memanfaatkan infrastruktur *cloud Google* dan memberikan lingkungan pengembangan yang kuat dengan akses

ke GPU dan TPU (unit pemrosesan tensor) secara gratis. *Google Colab* memanfaatkan infrastruktur *cloud Google* dan dapat diakses melalui peramban web, tidak memerlukan instalasi perangkat lunak tambahan, dan menyediakan kemudahan kolaborasi dengan memungkinkan pengguna untuk berbagi notebook secara langsung dengan orang lain [16]. Berikut ini adalah beberapa keuntungan dan kelebihan *Google Colab*.

- a. Menulis dan menjalankan kode *Python*
- b. Membuat, mengunggah, dan membagikan buku catatan
- c. Mengimpor atau menyimpan buku catatan dari atau ke *Google Drive*
- d. Mengimpor atau mempublikasikan buku catatan dari GitHub
- e. Mengimpor set data eksternal Mengintegrasikan *PyTorch*, *TensorFlow*, *OpenCV*
- f. Memberikan layanan *cloud* gratis dengan CPU gratis
- g. Memudahkan pengguna untuk menjalankan program pada komputer dengan spesifikasi tinggi.¹
- h. Lebih fleksibel khususnya dalam hal integrasi sebab terhubung dengan layanan *Google* yang lainnya, seperti *Google Drive*.

Google Colab saat ini banyak di minati khususnya bagi para pebisnis, programmer dan researcher, karena *Google Colab* sendiri memiliki manfaat dan fitur yang cukup lengkap. Fitur tersebut dapat membantu para pengguna untuk

mendapatkan fasilitas seperti layaknya menggunakan komputer dengan fitur yang lengkap secara gratis untuk menjalankan program yang telah dibuat olehnya.

2.2.8 Cara kerja DBSCAN

1. Pengumpulan data

Kumpulan data pengeluaran dari wisatawan toraja, misalnya:

- a. Akomodasi
- b. Makanan & minuman
- c. Transportasi
- d. Belanja

2. Pra-pemrosesan data

- a. Bersihkan data dari nilai kosong atau tidak logis
- b. Normalisasi/Standarisasi data agar semua fitur dalam skala yang sama.

3. Menentukan parameter DBSCAN

- a. *Epsilon*: Jarak maksimum antara wisatawan agar dianggap tetangga
- b. *MinPts*: Jumlah minimal tetangga dalam radius agar tidak dianggap pusat cluster.

4. Menghitung jarak

5. Identifikasi titik

- a. Jika dalam radius ϵ terdapat *MinPts* wisatawan lain→jadi core point
- b. Jika tidak, tapi berada dekat core poin→jadi border point
- c. Jika tidak masuk dua-duanya→jadi *noise/outlier*

6. Pembentukan cluster

- a. Hubungkan core points yang saling berdekatan menjadi satu cluster.

- b. *Border points* mengikuti *coree point* terdekat.
- c. *Noise* tidak masuk ke *cluster* manapun.

7. Hasil Clustering

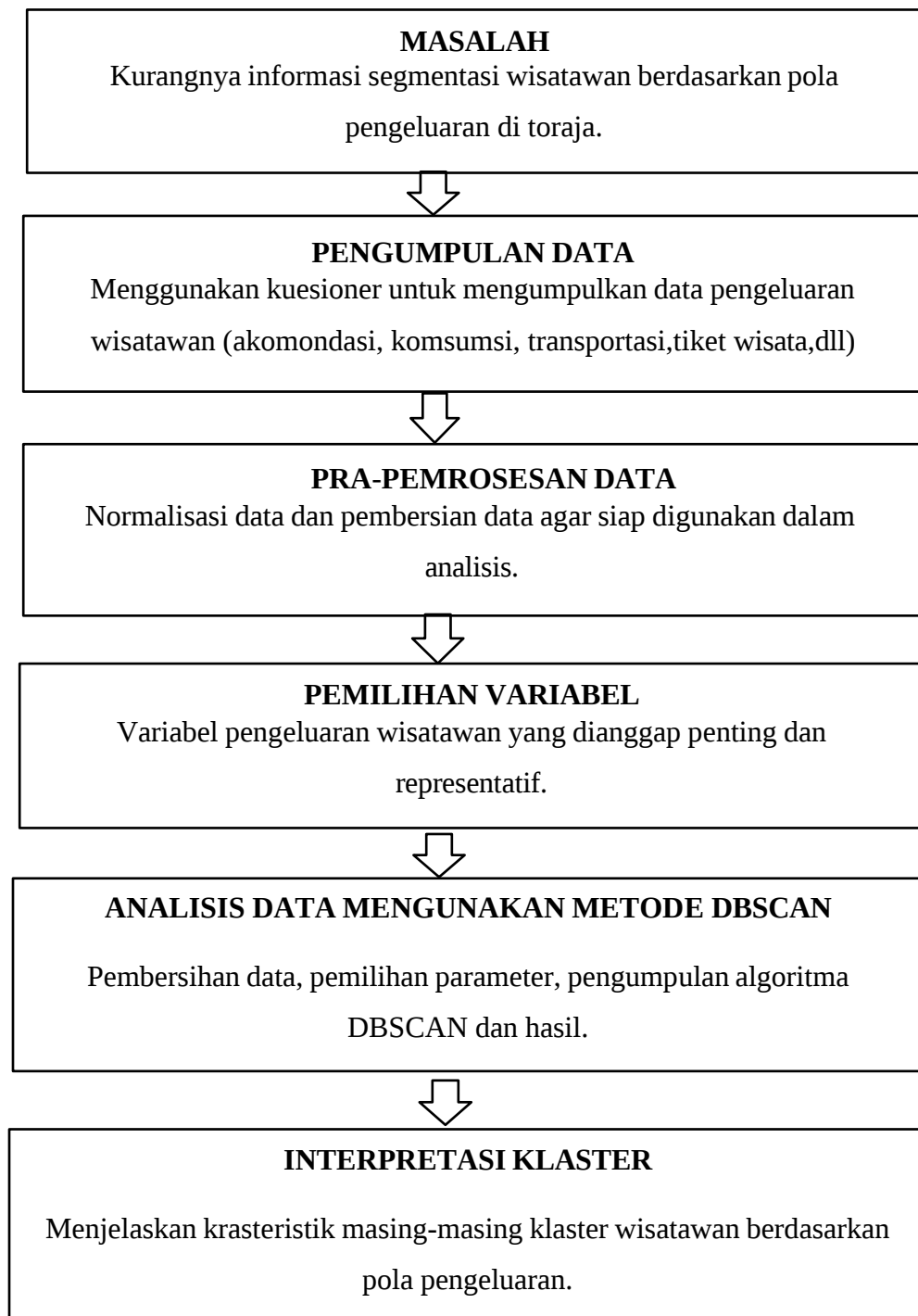
- a. Pengeluaran besar untuk akomondasi dan makanan.
- b. Hemat dan banyak belanja oleh-oleh.
- c. Fokus pada atraksi budaya dan transportasi.

8. Analisis dan pemanfaatan

- a. Menentukan segmentasi pasar wisatawan.
- b. Menyusun strategi promosi
- c. Mendesain paket wisata yang sesuai dengan perilaku pengeluaran.

2.3 Kerangka Pikir

Adapun kerangka pikir dari penelitian ini seperti pada gambar 3.1



Gambar 2. 2 Kerangka pikir