BAB III
METODE PENELITIAN

3.1 Objek Penelitian

Hardani dkk. (2020) menjelaskan bahwa objek penelitian merupakan target ilmiah yang digunakan untuk mengumpulkan data guna mencapai tujuan tertentu dan harus bersifat objektif, valid, dapat diandalkan, serta dapat berupa individu, benda, atau fenomena yang menjadi fokus utama dalam penelitian. Berdasarkan definisi tersebut, dapat disimpulkan bahwa objek penelitian adalah segala sesuatu yang menjadi pusat perhatian dalam suatu penelitian, baik berupa manusia, benda, maupun fenomena, guna memperoleh data yang relevan dengan tujuan penelitian.

Objek dalam penelitian ini adalah data profil dan karakteristik pengunjung *event* Vindes Bukan Main, yang meliputi variabel demografis, geografis, psikografis, dan perilaku.

3.2 Metode Penelitian

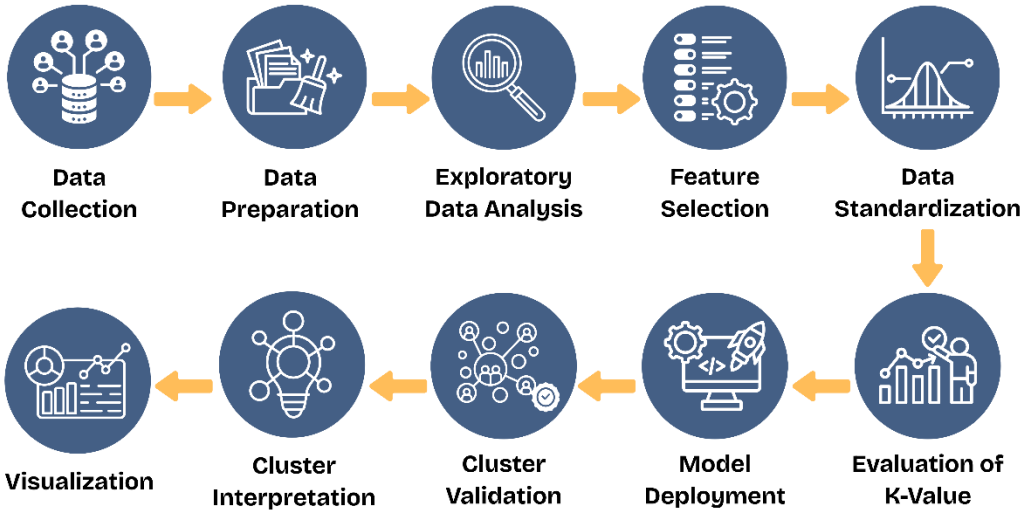
Metode penelitian merupakan cara atau pendekatan yang digunakan untuk mengumpulkan serta menganalisis data dalam suatu studi (Sarie dkk, 2022). Dalam penelitian ini, metode yang digunakan adalah *clustering*. Metode *clustering* merupakan pendekatan eksploratif yang bertujuan untuk mengelompokkan objek-objek ke dalam beberapa kelompok (*cluster*) berdasarkan tingkat kemiripan karakteristik antar objek, tanpa adanya label kelas yang diketahui sebelumnya. Teknik ini digunakan untuk menemukan struktur alami atau pola tersembunyi dalam sekumpulan data yang belum terklasifikasi, sehingga sangat cocok diterapkan dalam penelitian yang berfokus pada segmentasi atau pemetaan karakteristik objek berdasarkan data yang tersedia (Triayudi dkk., 2024).

Dalam konteks penelitian ini, metode *clustering* digunakan untuk mengelompokkan data pengunjung sebuah *event* berdasarkan variabel segmentasi. Tujuannya adalah untuk mengidentifikasi segmen-segmen pengunjung yang memiliki kesamaan profil, perilaku, preferensi, atau karakteristik tertentu. Dengan menggunakan metode ini, diharapkan penelitian mampu memberikan wawasan

strategis yang dapat dimanfaatkan untuk memahami dinamika pengunjung secara lebih terstruktur dan mendalam.

3.2.1 Desain Penelitian

Desain penelitian adalah rencana dan strategi yang disusun untuk menjawab rumusan masalah, sehingga proses penelitian dapat berjalan secara efektif, efisien, dan terarah (Pasaribu, Herawati, Utomo, & Aji, 2022). Dalam penelitian ini, desain penelitian terdiri dari tahapan-tahapan *clustering* yang mengacu pada model Ashwani dkk. (2023). Tahapan tersebut yang tertera pada Gambar 3.1 berikut ini.



Gambar 3.1 Tahapan K-Means *Clustering*

Berikut adalah penjelasan untuk setiap tahapan yang akan dilakukan.

1. *Data Collection*

   Langkah pertama pada metode ini adalah mengumpulkan data. Data pengunjung dikumpulkan melalui kuesioner yang disebarkan kepada pengunjung yang pernah menghadiri secara langsung acara "Vindes Bukan Main". Data yang dikumpulkan mencakup empat atribut utama berdasarkan teori segmentasi Kotler dkk. (2020), yaitu demografis, geografis, psikografis, dan perilaku.

   Namun, tidak semua indikator dalam masing-masing variabel segmentasi digunakan secara utuh sebagaimana dijelaskan dalam teori segmentasi tersebut. Penyesuaian dilakukan dengan mempertimbangkan relevansi terhadap konteks

industri *event*. Misalnya, pada variabel demografis, hanya atribut yang dianggap signifikan dalam memengaruhi perilaku pengunjung *event*, seperti usia, jenis kelamin, pekerjaan yang dimasukkan, sementara atribut lain seperti agama atau etnis tidak digunakan.

Selain melakukan penyesuaian dengan tidak menggunakan seluruh contoh indikator yang disebutkan dalam teori segmentasi, penelitian ini juga menambahkan beberapa indikator yang relevan dengan karakteristik pengunjung *event*. Penambahan ini tetap mengacu pada variabel segmentasi, tetapi disesuaikan dengan konteks kebutuhan di dunia *event*. Sebagai contoh, untuk variabel perilaku yang secara umum mencakup loyalitas, digunakan indikator seperti frekuensi kehadiran pengunjung pada *event* yang diselenggarakan oleh Vindes. Tabel 3.1 merincikan variabel, dimensi, dan indikator yang digunakan dalam kuesioner.

Tabel 3.1 Operasionalisasi Variabel

Variabel	Dimensi	Indikator
Demografis	Usia	Rentang usia
	Jenis kelamin	Kategori jenis kelamin
	Pekerjaan	Kategori pekerjaan
Geografis	Domisili	Nama kota tempat tinggal
Psikografis	Motivasi	Alasan yang mendorong kehadiran
	Preferensi	Kecenderungan diri dalam memilih hiburan
		Preferensi aktivitas saat menghadiri acara
		Jenis produk yang paling sering dibeli di Vindes Bukan Main
		Area favorit yang sering dikunjungi di Vindes Bukan Main
		Pendamping ke *event*
		Preferensi pengeluaran untuk hiburan per bulan
Perilaku	Loyalitas	Jumlah hari menghadiri Vindes Bukan Main

Variabel	Dimensi	Indikator
		Frekuensi ke *event* yang diselenggarakan Vindes
		Durasi waktu yang dihabiskan di Vindes Bukan Main
	Pengeluaran	Nominal uang yang dibelanjakan saat di Vindes Bukan Main
	Sumber informasi	Media atau cara pertama mengetahui acara
	Durasi waktu	Lama waktu pengunjung berada di acara Vindes Bukan Main
	Transportasi	Jenis moda transportasi yang digunakan

2. *Data Preparation*

Sebelum dianalisis, data harus dibersihkan dan diformat terlebih dahulu. Proses ini mencakup memastikan bahwa setiap variabel memiliki format yang sesuai, menghapus duplikasi, serta membersihkan data untuk memastikan keakuratan. *Outlier* atau nilai ekstrem juga perlu diidentifikasi dan ditangani, baik dengan menghapus maupun mentransformasikannya agar tidak memengaruhi hasil analisis. Sebagai bagian dari tahap persiapan data, atribut yang tidak relevan terlebih dahulu dihapus, lalu data diperiksa dan dibersihkan dari nilai kosong (*null*).

Dalam penerapan algoritma K-Means *clustering*, data yang digunakan harus dalam format numerik. Oleh karena itu, data kategorik perlu dikonversi menggunakan teknik *label encoding*, yaitu metode pengubahan nilai kategorik menjadi angka. Proses ini dilakukan dengan memberikan label berupa angka pada setiap kategori, dimulai dari angka 0 untuk menjaga indeks yang konsisten dan sesuai dengan praktik umum pemrograman serta komputasi numerik (McKinney, 2022).

3. *Exploratory Data Analysis* (EDA)

Disarankan untuk melakukan *Exploratory Data Analysis* (EDA) terlebih dahulu guna memahami karakteristik data sebelum menerapkan algoritma K-

Means *clustering*. Dalam proses ini, pembuatan ringkasan statistik dan visualisasi data sangat membantu untuk mengenali pola, struktur, serta hubungan tersembunyi dalam data.

4. *Feature Selection*

   Langkah selanjutnya adalah memilih variabel yang akan digunakan dalam analisis. Pemilihan ini penting karena algoritma K-Means bekerja lebih optimal jika variabel yang digunakan tidak memiliki korelasi tinggi satu sama lain. Oleh karena itu, perlu dipilih sejumlah variabel yang paling relevan dengan tujuan penelitian dan saling tidak berkorelasi tinggi. Pada tahap ini, peneliti berfokus pada variabel-variabel yang dianggap signifikan untuk proses penerapan K-Means *clustering*. *Feature selection* ini juga bertujuan untuk mengurangi *curse of dimensionality*, yaitu kurangnya performa model yang disebabkan oleh penggunaan banyak dimensi.

   Menurut Schober, Boer, dan Schwarte (2018), interpretasi nilai koefisien korelasi dapat diklasifikasikan sebagai berikut: 0,00-0,10 menunjukkan korelasi sangat lemah, 0,10-0,39 menunjukkan korelasi lemah, 0,40-0,69 menunjukkan korelasi sedang, 0,70-0,89 menunjukkan korelasi kuat, dan 0,90-1,00 menunjukkan korelasi sangat kuat. Sementara itu, nilai -1 menunjukkan adanya korelasi negatif sempurna antara dua variabel.

5. *Data Standardization*

   Standardisasi atau normalisasi *Z-score* dilakukan untuk mengubah skala data sehingga memiliki nilai rata-rata 0 dan standar deviasi 1. Karena fitur memiliki skala berbeda, ada risiko fitur dengan nilai lebih besar mendapat bobot lebih tinggi, yang dapat memengaruhi kinerja algoritma. Untuk menghindari bias tersebut, khususnya pada algoritma berbasis jarak, data perlu diskalakan agar setiap fitur memberikan kontribusi yang setara terhadap hasil analisis. *Data standardization* pada Python menggunakan fungsi *StandardScaler* (Vinai, 2021).

6. *Evaluation of* K-*value*

   Penentuan jumlah *cluster* (K) optimal akan menggunakan metode *Elbow*. Metode ini mengevaluasi variasi dalam *cluster* (*Within-Cluster Sum of Squares/WCSS*) terhadap jumlah *cluster* dan membantu menemukan titik "*elbow*" dimana penurunan WCSS mulai melambat, menunjukkan jumlah *cluster* yang tepat untuk segmentasi yang optimal. Namun, *Elbow Method* hanya memberikan titik awal yang baik untuk menentukan nilai K. Pemilihan akhir tetap perlu dievaluasi dan divalidasi menggunakan metriks yang lain.

7. *Model Deployment*

   Setelah nilai K ditentukan, algoritma K-Means *clustering* diterapkan pada *dataset* untuk membentuk *cluster* berdasarkan fitur terpilih. Pada tahap ini dilakukan juga pemodelan dengan K-1 dan K+1 sebagai perbandingan.

8. *Cluster Validation*

   Validasi kualitas *cluster* dilakukan menggunakan *Silhouette Score*. Metode ini mengukur seberapa baik data dikelompokkan dengan mempertimbangkan jarak antar titik data dalam *cluster* yang sama dan jarak dengan *cluster* lain. Nilai *Silhouette* yang tinggi menandakan *cluster* yang rapat dan terpisah dengan baik, sehingga segmentasi dapat dipercaya.

9. *Cluster Interpretation*

   Tahap interpretasi dilakukan dengan memberi label pada setiap *cluster* berdasarkan ciri khasnya. Profil klaster disusun untuk memahami gambaran umum, menilai kesesuaian anggota *cluster*, dan memastikan tiap klaster memiliki karakteristik yang berbeda secara jelas (Muttaqin, 2022).

10. *Visualization of* K-Means *Clustering*

    Visualisasi hasil *clustering* dapat dilakukan menggunakan *scatter plot* 2D dan 3D untuk memberikan gambaran yang jelas mengenai distribusi *cluster* dalam ruang fitur yang dipilih, sehingga memudahkan pemahaman dan penyampaian hasil kepada pihak terkait.

3.2.2 Jenis dan Sumber Data

Penelitian ini menggunakan jenis data kuantitatif sebagai sumber data utama, yang menurut Sarie dkk. (2022) berkaitan dengan angka dalam pengumpulan dan pengolahan data. Selain itu, ata kualitatif juga digunakan sebagai data pendukung untuk memperkaya pemahaman terhadap fenomena yang diteliti. Data yang digunakan terdiri atas data primer dan sekunder. Data primer diperoleh melalui wawancara dengan pihak internal Vindes serta penyebaran kuesioner kepada pengunjung *event* Vindes Bukan Main. Sementara itu, data sekunder berasal dari literatur yang relevan, seperti teori segmentasi, algoritma K-Means *clustering*, *event* Vindes Bukan Main, *data-driven marketing*, dan visualisasi data.

3.2.3 Populasi dan Sampel

Amin, Garancang, dan Abunawas (2023) menyatakan bahwa populasi adalah seluruh komponen dalam penelitian mencakup objek dan subjek yang memiliki ciri serta karakteristik khusus. Berdasarkan definisi tersebut, maka populasi dalam penelitian ini adalah seluruh pengunjung *event* Vindes Bukan Main.

Sampel merupakan bagian terbatas dari populasi yang dianggap dapat merepresentasikan keseluruhan populasi (Ramdani & Utami, 2022). Sampel pada penelitian ini adalah pengunjung Vindes Bukan Main yang benar-benar hadir saat acara berlangsung dan bersedia mengisi kuesioner yang disebarkan secara daring.

3.2.4 Teknik Penarikan Sampel

Pemilihan sampel dilakukan menggunakan teknik *non-probability sampling* melalui pendekatan *purposive sampling*. Metode ini dipilih karena sampel ditentukan berdasarkan kriteria tertentu, bahwa responden merupakan pengunjung yang telah menghadiri *event* Vindes Bukan Main secara langsung, yang diverifikasi melalui pertanyaan validasi kehadiran dalam kuesioner. Sedangkan untuk wawancara, sampel ditentukan dengan pendekatan *purposive sampling* berbasis informan kunci (*key informant sampling*), yaitu dua orang dari tim *Marketing & Communication* Vindes Bukan Main yang dianggap memiliki informasi relevan mengenai penyelenggaraan *event*. Dengan demikian, data yang diperoleh

diharapkan dapat merepresentasikan karakteristik populasi yang diteliti secara tepat.

Dalam penelitian berbasis *data science*, penentuan jumlah sampel merupakan tahap penting untuk memperoleh data yang representatif. Menurut Ramdani dan Utami (2022), terdapat beberapa teknik penarikan sampel yang dapat digunakan dalam *data science*, salah satunya adalah dengan menggunakan rumus Yamane.

Rumus Yamane pertama kali diperkenalkan oleh Yamane (1967) untuk menentukan jumlah sampel dari populasi yang diketahui, dengan persamaan sebagai berikut.

$$n = \frac{N}{1 + N(e)^2}$$

Dengan:

n : jumlah sampel

N : jumlah populasi

e : tingkat presisi yang ditetapkan peneliti

Diketahui jumlah populasi pengunjung *event* Vindes Bukan Main adalah 3000 pengunjung dan tingkat presisi yang ditetapkan adalah 0.10 atau setara dengan tingkat kepercayaan 90%. Dengan demikian, jumlah sampel yang harus dikumpulkan dengan rumus tersebut adalah sebagai berikut.

$$n = \frac{3000}{1 + 3000(0.1)^2} = \frac{3000}{1 + 3000(0.01)} = \frac{3000}{1 + 30} = \frac{3000}{31} \approx 96.77 \approx 100$$

Berdasarkan perhitungan rumus di atas, maka jumlah sampel yang harus dikumpulkan adalah sebanyak 96 orang dan dibulatkan menjadi 100 orang.

3.2.5 Teknik Pengumpulan Data

Teknik pengumpulan data merupakan metode yang digunakan oleh untuk memperoleh data yang dibutuhkan dalam penelitian (Pasaribu dkk., 2022). Dalam penelitian ini, teknik pengumpulan data yang digunakan meliputi wawancara, studi literatur, dan kuesioner.

A. Wawancara

Wawancara merupakan salah satu teknik pengumpulan data yang digunakan dalam penelitian ini untuk memperoleh informasi terkait data jumlah

pengunjung, pemanfaatan data pengunjung, dan validasi hasil segmentasi. Teknik ini dipilih karena wawancara memungkinkan peneliti untuk mendapatkan data kualitatif yang lebih kaya, serta menggali perspektif langsung dari pihak yang memiliki keterlibatan dalam operasional dan strategi bisnis Vindes. Dalam penelitian ini, wawancara dilakukan dengan dua orang internal Vindes yang memiliki wawasan mendalam mengenai penyelenggaraan *event* Vindes terutama Vindes Bukan Main.

B. Studi Literatur

Studi literatur dilakukan dengan mengumpulkan data sekunder dari sumber-sumber kredibel seperti buku, *e-book*, jurnal, dan *website* terpercaya. Referensi yang dicari terkait topik *clustering*, algoritma K-Means, segmentasi, serta informasi tentang *event* Vindes Bukan Main. Studi literatur ini bertujuan untuk memperkuat landasan teoritis dan konseptual penelitian.

C. Kuesioner

Kuesioner digunakan sebagai teknik pengumpulan data primer dalam penelitian ini. Penyebaran dilakukan secara daring melalui *platform* Google Form, yang dibagikan kepada pengunjung *event* Vindes Bukan Main melalui media sosial. Instrumen kuesioner disusun berdasarkan teori segmentasi, di mana setiap indikator pertanyaan dirancang sesuai dengan dimensi dalam teori tersebut dan disesuaikan dengan konteks *event*. Jenis kuesioner yang digunakan adalah kuesioner tertutup, sehingga responden hanya perlu memilih jawaban yang telah disediakan, guna mempermudah proses pengisian dan analisis data.

3.2.6 Teknik Analisis Data

Menurut Wada dkk. (2024), teknik analisis data merupakan pendekatan yang digunakan untuk mengelola, menganalisis, serta menyajikan data guna memperoleh makna atau kesimpulan dari hasil penelitian. Berikut teknik analisis data yang digunakan pada penelitian ini.

A. Analisis Deskriptif

Analisis deskriptif digunakan untuk menggambarkan atau mendeskripsikan data yang telah dikumpulkan dalam penelitian. Teknik ini bertujuan untuk

memberikan gambaran umum tentang karakteristik data, seperti modus dan distribusi frekuensi atau proporsi kategori. Dalam konteks penelitian ini, analisis deskriptif digunakan dalam tahap *Exploratory Data Analysis* (EDA) dan untuk memaparkan profil pengunjung *event* Vindes Bukan Main berdasarkan variabel-variabel yang diamati.

B. *Exploratory Data Analysis* (EDA)

Exploratory Data Analysis (EDA) adalah teknik analisis data yang bertujuan untuk mengeksplorasi data secara mendalam guna mengidentifikasi pola, distribusi, dan korelasi dalam data. EDA melibatkan penggunaan visualisasi data, seperti *bar chart*, *pie chart*, dan *correlation heatmap* untuk memahami proporsi data dan hubungan antar variabel.

C. *Clustering*

Clustering adalah teknik analisis data yang digunakan untuk mengelompokkan data ke dalam beberapa kelompok (*cluster*) berdasarkan kesamaan karakteristik (Muhima dkk., 2021). Dalam penelitian ini, *clustering* akan dilakukan menggunakan algoritma K-Means. Algoritma K-Means bekerja dengan mengelompokkan data ke dalam sejumlah *cluster* yang telah ditentukan (nilai K) berdasarkan jarak *Euclidean* antara titik data dan *centroid cluster*. Melalui teknik *clustering*, penelitian ini dapat mengidentifikasi segmentasi pengunjung *event* Vindes Bukan Main berdasarkan kesamaan karakteristik, sehingga dapat memberikan rekomendasi yang lebih tepat untuk meningkatkan pengalaman pengunjung.

3.3 Alat dan Bahan Penelitian

3.3.1 Alat

Berikut spesifikasi perangkat keras (*hardware*) dan perangkat lunak (*software*) yang digunakan pada penelitian ini.

1. Perangkat Keras

   Satu unit laptop dengan spesifikasi:

Tabel 3.2 Spesifikasi Perangkat Keras

No.	Nama	Spesifikasi
1	Processor	AMD Ryzen 5 5600U with Radeon Graphics

No.	Nama	Spesifikasi
2	Graphic	NVIDIA GeForce RTX 3050
3	RAM	16 GB
4	Sistem Operasi	Windows 11 64-bit

2. Perangkat Lunak

   A. Google Colab

      Google Colab adalah *platform* berbasis *cloud* yang memungkinkan pengguna menjalankan kode Python secara interaktif tanpa perlu menginstal perangkat lunak tambahan di komputer lokal. Google Colab mendukung penggunaan pustaka *machine learning* dan *data science* seperti *NumPy*, *pandas*, dan *Scikit-Learn*, sehingga sangat cocok untuk implementasi algoritma K-Means *clustering* dalam penelitian ini.

3.3.2 Bahan

Bahan utama dalam penelitian ini adalah data pengunjung *event* Vindes Bukan Main yang diperoleh melalui kuesioner daring. Kuesioner disusun dalam format Google Form dan disebarkan melalui media sosial untuk menjangkau audiens yang berpartisipasi dalam *event* tersebut.