

BAB II

LANDASAN TEORI

2.1 Penelitian Terdahulu

Tabel 2.1 Penelitian Terdahulu

No	Peneliti & Tahun	Metode Penelitian	Sitasi	Kesimpulan
1	Putri, Safwandi, & Fitri (2025)	K-Means Clustering	17	K-Means berhasil mengelompokkan pelanggan ke dalam beberapa segmen berdasarkan karakteristik yang dimiliki sehingga membantu proses pengambilan keputusan bisnis.
2	Saputra & Yuliana (2023)	CRISP-DM	21	Framework CRISP-DM mampu memberikan tahapan yang sistematis dalam proses data mining mulai dari pemahaman bisnis hingga deployment.
3	Kurniawan & Aditya (2024)	K-Means Clustering	22	K-Means efektif digunakan untuk mengelompokkan data berdasarkan kemiripan karakteristik melalui proses perhitungan centroid.
4	Andini, Liantoni, & Budianto (2025)	K-Means Clustering	31	Algoritma K-Means berhasil mengelompokkan kompetensi mahasiswa ke dalam beberapa cluster

				yang memiliki karakteristik berbeda.
5	Johan & William (2025)	K-Means Clustering	34	Segmentasi pelanggan menggunakan K-Means menghasilkan kelompok pelanggan yang dapat digunakan untuk menentukan strategi pemasaran yang lebih tepat.
6	Saputra (2024)	Clustering (K-Means)	35	K-Means mampu mengidentifikasi segmen konsumen berdasarkan data survei sehingga membantu memahami karakteristik masing-masing kelompok konsumen.
7	Hendrastuty et al. (2024)	Clustering (K-Means)	36	K-Means dapat digunakan untuk menemukan pola dan karakteristik kelompok yang tersembunyi dalam data sehingga menghasilkan insight untuk pengambilan keputusan.
8	Muttaqin et al. (2022)	CRISP-DM dan K-Means	37	Kombinasi CRISP-DM dan K-Means mampu menghasilkan segmentasi data yang terstruktur dan mendukung proses analisis yang lebih sistematis.
9	Hayati, K. R., Nurhayati, S., & Prasetyo, A. (2024).	Uji Validalitas (Pearson)	38	Penelitian menunjukkan bahwa Pearson Product Moment Correlation dapat digunakan untuk mengukur validitas setiap item pertanyaan dengan membandingkan nilai

				korelasi item terhadap total skor. Item dinyatakan valid apabila nilai korelasi yang diperoleh memenuhi kriteria yang telah ditentukan.
10	Ahmad, N., Alias, F. A., Hamat, M., & Mohamed, S. A. (2024).	Reliabilitas (Cronbach Alpha)	39	Cronbach's Alpha merupakan metode yang paling banyak digunakan untuk mengukur konsistensi internal instrumen penelitian. Nilai alpha yang tinggi menunjukkan bahwa item-item pertanyaan mampu mengukur konstruk yang sama secara konsisten.
11	Koo, M. (2025).	Likert Scale	40	Skala Likert merupakan metode pengukuran yang banyak digunakan dalam penelitian survei untuk mengukur sikap, persepsi, dan tingkat persetujuan responden terhadap suatu pernyataan.

Pembahasan Penelitian terdahulu

Penelitian terdahulu menunjukkan bahwa algoritma K-Means Clustering telah banyak digunakan untuk melakukan segmentasi data pada berbagai bidang penelitian. Putri et al. [17] menerapkan K-Means Clustering untuk segmentasi pelanggan dan berhasil mengelompokkan pelanggan berdasarkan karakteristik yang dimiliki sehingga dapat mendukung pengambilan keputusan bisnis. Penelitian serupa juga dilakukan oleh Johan dan William [34] yang menggunakan K-Means untuk segmentasi pelanggan sebagai dasar dalam penyusunan strategi pemasaran yang lebih tepat sasaran.

Pada bidang survei dan karakteristik responden, Saputra [35] serta Andini et al. [31] menunjukkan bahwa algoritma K-Means mampu mengelompokkan responden berdasarkan kemiripan karakteristik sehingga menghasilkan informasi yang lebih terstruktur mengenai profil masing-masing kelompok. Hasil penelitian tersebut menunjukkan bahwa metode clustering dapat digunakan untuk mengidentifikasi pola tersembunyi yang sulit diamati melalui analisis deskriptif konvensional.

Selain penggunaan algoritma K-Means, beberapa penelitian juga menerapkan framework CRISP-DM sebagai metodologi dalam proses data mining. Saputra dan Yuliana [21] menyatakan bahwa CRISP-DM mampu memberikan tahapan kerja yang sistematis mulai dari business understanding hingga deployment. Temuan tersebut diperkuat oleh penelitian Muttaqin et al. [37] yang mengkombinasikan CRISP-DM dengan algoritma K-Means untuk menghasilkan proses segmentasi data yang lebih terstruktur dan mendukung pengambilan keputusan berbasis data.

Berdasarkan penelitian-penelitian terdahulu tersebut, dapat disimpulkan bahwa K-Means Clustering merupakan metode yang efektif untuk melakukan segmentasi data berdasarkan kemiripan karakteristik objek penelitian, sedangkan CRISP-DM dapat digunakan sebagai framework yang sistematis dalam proses analisis data. Namun, penelitian yang secara khusus membahas persebaran minat dan karakteristik masyarakat terhadap olahraga padel di wilayah Jabodetabek masih sangat terbatas. Oleh karena itu, penelitian ini dilakukan dengan menerapkan pendekatan data mining berbasis CRISP-DM dan algoritma K-Means Clustering untuk mengidentifikasi segmentasi masyarakat berdasarkan karakteristik demografis, aktivitas bermain, tingkat minat, serta faktor-faktor yang mempengaruhi ketertarikan terhadap olahraga padel menggunakan data primer dari 1.012 responden di wilayah Jabodetabek.

2.2 Teori yang berkaitan

2.2.1 Big Data Analytics

Big Data Analytics merupakan pendekatan yang digunakan untuk mengelola dan menganalisis data dalam skala besar dengan tingkat kompleksitas dan keragaman yang tinggi sehingga dapat menghasilkan informasi yang bermanfaat bagi proses pengambilan keputusan [10]. Konsep big data umumnya dijelaskan melalui lima karakteristik utama yang dikenal sebagai 5V, yaitu Volume, Velocity, Variety, Veracity, dan Value [19]. Kelima karakteristik tersebut menggambarkan besarnya jumlah data yang tersedia, tingginya kecepatan pertumbuhan data, beragamnya format data yang dihasilkan, perbedaan tingkat keakuratan data, serta potensi nilai yang dapat diperoleh dari pemanfaatan data tersebut.

Dalam bidang Sistem Informasi, Big Data Analytics memiliki peran penting dalam mengubah data mentah menjadi informasi yang lebih bermakna dan dapat digunakan sebagai dasar pengambilan keputusan [20]. Pemanfaatannya tidak hanya terbatas pada kegiatan prediksi, tetapi juga mencakup analisis deskriptif dan eksploratif untuk mengidentifikasi pola, kecenderungan, maupun karakteristik tertentu yang terdapat dalam data.

Pada penelitian ini, pendekatan Big Data Analytics diterapkan untuk mengolah data hasil kuesioner yang diperoleh melalui Google Form. Melalui proses analisis tersebut, peneliti dapat memperoleh pemahaman yang lebih mendalam mengenai karakteristik responden dan faktor-faktor yang mempengaruhi ketertarikan masyarakat terhadap olahraga padel.

Dalam praktiknya, penerapan Big Data Analytics sering dipadukan dengan berbagai teknik data mining dan machine learning untuk menghasilkan analisis yang lebih komprehensif [12]. Kombinasi metode tersebut memungkinkan proses identifikasi pola, pengelompokan data, klasifikasi, prediksi, hingga penemuan hubungan antarvariabel dilakukan

secara lebih efektif. Dengan demikian, organisasi maupun peneliti dapat memperoleh insight yang lebih akurat dan relevan untuk mendukung pengambilan keputusan yang berbasis data.

2.2.2 Visualisasi Data

Visualisasi data merupakan proses penyajian data dalam bentuk grafik, diagram, dashboard, maupun bentuk visual lainnya untuk mempermudah proses analisis dan interpretasi informasi [11]. Visualisasi data memiliki peran penting dalam proses Big Data Analytics karena mampu membantu pengguna memahami pola, tren, dan hubungan antar data secara lebih efektif.

Dalam penelitian ini, visualisasi data digunakan untuk menampilkan hasil analisis sentimen, persebaran minat masyarakat, hasil clustering, classification, serta Association Rule Mining terhadap olahraga padel di wilayah Jabodetabek. Visualisasi dilakukan menggunakan Power BI untuk menghasilkan dashboard interaktif yang mampu membantu proses interpretasi data dan penyampaian insight penelitian [20].

Visualisasi data juga membantu proses penyampaian hasil penelitian menjadi lebih informatif, mudah dipahami, dan mendukung pengambilan keputusan berbasis data.

2.2.3 Kuesioner dan Google Form sebagai Sumber Data

Kuesioner merupakan metode pengumpulan data yang digunakan untuk memperoleh informasi dari responden melalui serangkaian pertanyaan yang telah disusun secara sistematis [26]. Kuesioner sering digunakan untuk mengukur persepsi, sikap, dan minat masyarakat terhadap suatu fenomena.

Google Form merupakan salah satu media digital yang banyak digunakan untuk penyebaran kuesioner karena kemudahan akses, efisiensi waktu, serta kemampuan dalam mengelola dan merekap data secara otomatis [27]. Penggunaan Google Form memungkinkan data

dikumpulkan secara terstruktur dan siap diolah lebih lanjut menggunakan alat analisis data.

Dalam penelitian ini, Google Form digunakan sebagai sumber data utama untuk mengumpulkan informasi terkait minat dan sentimen masyarakat terhadap olahraga padel di wilayah Jabodetabek [11].

2.2.4 Bahasa Pemrograman Python dalam Analisis Data

Python merupakan salah satu bahasa pemrograman yang banyak dimanfaatkan dalam bidang analisis data, data science, dan Big Data Analytics. Popularitas Python didukung oleh sintaks yang mudah dipahami serta ketersediaan berbagai pustaka yang mampu mendukung proses pengolahan dan analisis data secara efektif [28]. Dengan kemampuannya tersebut, Python memungkinkan berbagai tahapan analisis data dilakukan secara lebih terstruktur, mulai dari pengolahan data hingga penyajian hasil dalam bentuk visualisasi yang informatif [29].

Pada penelitian ini, Python digunakan sebagai alat utama dalam mengolah data hasil kuesioner yang diperoleh dari responden. Selain mendukung proses analisis deskriptif dan analisis sentimen, Python juga dimanfaatkan untuk membuat berbagai bentuk visualisasi data yang bertujuan mempermudah interpretasi hasil penelitian. Penggunaan Python dalam penelitian ini mendukung penerapan konsep Big Data Analytics untuk menghasilkan informasi yang lebih akurat, relevan, dan mudah dipahami oleh pengguna [10].

Selain digunakan dalam pengolahan data, Python juga menjadi salah satu platform yang banyak diterapkan dalam pengembangan metode machine learning dan data mining. Hal tersebut didukung oleh berbagai pustaka yang tersedia, seperti Scikit-learn, Pandas, NumPy, dan Mlxtend, yang menyediakan beragam fungsi untuk melakukan proses clustering, classification, maupun association rule mining [29]. Ketersediaan pustaka

tersebut menjadikan Python sebagai salah satu bahasa pemrograman yang banyak dipilih dalam berbagai penelitian berbasis analisis data.

2.2.5 Data Mining

Data mining merupakan suatu proses analisis yang bertujuan untuk menggali pola, hubungan, serta informasi yang sebelumnya tidak terlihat secara langsung dari kumpulan data berukuran besar [21]. Proses ini memanfaatkan berbagai pendekatan, seperti teknik statistik, machine learning, dan pengelolaan basis data untuk menghasilkan pengetahuan yang bernilai dari data yang tersedia. Informasi yang diperoleh kemudian dapat digunakan sebagai dasar dalam mendukung pengambilan keputusan yang lebih efektif dan berbasis data.

Berbagai metode dapat diterapkan dalam proses data mining, di antaranya clustering, classification, regression, dan association rule mining [12]. Setiap metode memiliki fungsi yang berbeda sesuai dengan tujuan analisis yang dilakukan. Penerapan metode-metode tersebut memungkinkan peneliti untuk mengidentifikasi karakteristik data, menemukan keterkaitan antarvariabel, serta mengungkap pola yang dapat memberikan pemahaman lebih mendalam terhadap suatu fenomena.

Dalam penelitian ini, teknik data mining diterapkan untuk menganalisis data survei yang berkaitan dengan minat masyarakat terhadap olahraga padel. Proses analisis dilakukan menggunakan kombinasi metode K-Means Clustering, Classification, dan Association Rule Mining guna memperoleh gambaran yang lebih komprehensif mengenai karakteristik responden, tingkat minat masyarakat, serta faktor-faktor yang berhubungan dengan perkembangan olahraga padel.

2.2.6 Framework CRISP-DM

Cross Industry Standard Process for Data Mining (CRISP-DM) merupakan salah satu framework yang banyak digunakan dalam penelitian berbasis data mining dan data science karena menyediakan alur

kerja yang sistematis dalam proses analisis data [13]. Framework ini dirancang untuk membantu peneliti maupun praktisi dalam mengelola seluruh tahapan proyek data secara terstruktur, mulai dari perumusan masalah hingga penyajian hasil analisis.

CRISP-DM terdiri atas enam tahapan utama, yaitu Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, dan Deployment [22]. Setiap tahapan memiliki peran yang saling berkaitan untuk memastikan proses pengolahan dan analisis data dapat berjalan secara terarah serta menghasilkan temuan yang sesuai dengan tujuan penelitian.

Tahap pertama, Business Understanding, berfokus pada identifikasi permasalahan, tujuan, dan kebutuhan yang ingin dicapai dalam penelitian. Setelah itu, Data Understanding dilakukan untuk mempelajari karakteristik data, mengenali atribut yang tersedia, serta memahami kualitas data yang akan dianalisis. Selanjutnya, pada tahap Data Preparation dilakukan proses pembersihan, transformasi, dan penyiapan data agar siap digunakan pada tahap analisis.

Tahap Modeling merupakan proses penerapan metode atau algoritma yang dipilih untuk menjawab tujuan penelitian. Dalam penelitian ini, tahap tersebut mencakup penggunaan metode K-Means Clustering, Classification, dan Association Rule Mining. Hasil yang diperoleh kemudian dianalisis pada tahap Evaluation untuk menilai kesesuaian model dan memastikan bahwa hasil analisis telah memenuhi kebutuhan penelitian. Tahap terakhir, yaitu Deployment, dilakukan dengan menyajikan hasil analisis dalam bentuk informasi, visualisasi, maupun insight yang dapat dimanfaatkan sebagai dasar pengambilan keputusan [22].

CRISP-DM menjadi salah satu metodologi yang banyak diadopsi dalam berbagai penelitian karena memiliki struktur yang jelas, mudah dipahami,

serta fleksibel untuk diterapkan pada berbagai jenis permasalahan dan karakteristik data. Keunggulan tersebut menjadikan CRISP-DM sebagai framework yang relevan dalam mendukung proses analisis data secara sistematis dan berorientasi pada tujuan penelitian [22].

2.2.7 K-Means Clustering

K-Means Clustering merupakan salah satu algoritma dalam kategori unsupervised learning yang digunakan untuk mengelompokkan data berdasarkan tingkat kemiripan karakteristik yang dimiliki oleh setiap objek [14]. Tujuan utama dari metode ini adalah membentuk kelompok data yang memiliki karakteristik serupa dalam satu cluster, sementara perbedaan antar cluster dibuat semaksimal mungkin. Proses pengelompokan dilakukan dengan memanfaatkan jarak antara data dan pusat cluster (centroid) sebagai dasar penentuan keanggotaan suatu data [31].

Metode K-Means banyak diterapkan dalam berbagai penelitian karena memiliki mekanisme yang relatif sederhana, mudah dipahami, serta mampu mengolah data numerik dalam jumlah besar secara efisien. Pada prosesnya, algoritma akan menentukan sejumlah centroid awal yang kemudian digunakan sebagai acuan untuk mengelompokkan data. Setiap data akan ditempatkan pada cluster dengan centroid terdekat, dan posisi centroid akan diperbarui secara iteratif hingga diperoleh hasil pengelompokan yang stabil [23].

Dalam penelitian ini, K-Means Clustering digunakan untuk mengidentifikasi kelompok responden yang memiliki karakteristik serupa berdasarkan tingkat minat terhadap olahraga padel, frekuensi bermain, serta berbagai aktivitas yang berkaitan dengan olahraga tersebut. Hasil pengelompokan ini diharapkan dapat memberikan gambaran mengenai pola perilaku dan karakteristik masyarakat terhadap perkembangan olahraga padel di wilayah Jabodetabek.

2.2.8 Skala Likert

Skala Likert merupakan salah satu metode pengukuran yang paling banyak digunakan dalam penelitian survei untuk mengukur sikap, persepsi, opini, maupun tingkat persetujuan responden terhadap suatu pernyataan. Pada skala ini, responden diminta untuk memberikan penilaian terhadap suatu pernyataan berdasarkan tingkat kesetujuan atau tingkat preferensi yang telah ditentukan oleh peneliti. Data yang dihasilkan dari skala Likert termasuk ke dalam data ordinal karena menunjukkan urutan tingkat penilaian dari rendah hingga tinggi [40].

Menurut Koo [40], skala Likert memungkinkan peneliti mengukur konstruk yang bersifat subjektif secara lebih terstruktur dan kuantitatif. Penggunaan skala Likert banyak diterapkan dalam penelitian sosial, perilaku konsumen, pendidikan, maupun survei masyarakat karena mudah dipahami oleh responden serta mampu menghasilkan data yang dapat dianalisis secara statistik.

Pada penelitian ini, skala Likert digunakan untuk mengukur tingkat minat masyarakat terhadap olahraga padel, pengaruh media sosial, persepsi terhadap perkembangan olahraga padel, serta faktor-faktor yang mempengaruhi ketertarikan responden terhadap olahraga tersebut. Nilai yang diperoleh dari jawaban responden selanjutnya dikonversi ke dalam bentuk numerik untuk mendukung proses analisis data dan penerapan algoritma K-Means Clustering.

2.2.9 Uji Validitas

Uji validitas merupakan proses pengujian yang dilakukan untuk mengetahui sejauh mana instrumen penelitian mampu mengukur apa yang seharusnya diukur. Suatu instrumen dikatakan valid apabila setiap item pertanyaan dapat merepresentasikan variabel atau konstruk yang ingin diteliti secara tepat [38].

Menurut Wibowo dan Nugroho [38], pengujian validitas pada data survei dapat dilakukan menggunakan korelasi Pearson Product Moment dengan cara mengukur hubungan antara skor setiap item pertanyaan terhadap skor total variabel. Nilai koefisien korelasi yang diperoleh kemudian dibandingkan dengan nilai kritis yang telah ditentukan. Item pertanyaan dinyatakan valid apabila nilai korelasinya lebih besar dari nilai r tabel atau memiliki nilai signifikansi kurang dari 0,05.

Dalam penelitian ini, uji validitas dilakukan terhadap item-item pertanyaan yang menggunakan skala Likert untuk memastikan bahwa setiap pertanyaan mampu mengukur aspek yang berkaitan dengan minat, aktivitas bermain, serta faktor-faktor yang mempengaruhi ketertarikan masyarakat terhadap olahraga padel di wilayah Jabodetabek. Pengujian validitas dilakukan menggunakan metode Pearson Correlation dengan bantuan bahasa pemrograman Python.

2.2.10 Uji Reliabilitas

Uji reliabilitas merupakan pengujian yang bertujuan untuk mengetahui tingkat konsistensi suatu instrumen penelitian dalam mengukur variabel yang sama pada kondisi yang berbeda. Instrumen yang reliabel akan menghasilkan data yang relatif konsisten apabila digunakan berulang kali pada objek penelitian yang memiliki karakteristik serupa [39].

Salah satu metode yang paling umum digunakan dalam pengujian reliabilitas adalah Cronbach's Alpha. Menurut Taber [39], nilai Cronbach's Alpha digunakan untuk mengukur tingkat konsistensi internal antar item pertanyaan dalam suatu variabel. Semakin tinggi nilai Cronbach's Alpha, semakin baik tingkat reliabilitas instrumen penelitian. Secara umum, nilai Cronbach's Alpha $\geq 0,70$ menunjukkan bahwa instrumen memiliki tingkat reliabilitas yang baik dan dapat digunakan dalam proses analisis lebih lanjut.

Pada penelitian ini, uji reliabilitas dilakukan terhadap seluruh item pertanyaan yang menggunakan skala Likert untuk memastikan konsistensi jawaban responden terhadap variabel yang diukur. Pengujian dilakukan menggunakan metode Cronbach's Alpha dengan bantuan bahasa pemrograman Python sebelum data digunakan pada tahap pemodelan menggunakan algoritma K-Means Clustering.

2.2.11 Principal Component Analysis (PCA)

Principal Component Analysis (PCA) merupakan teknik reduksi dimensi yang digunakan untuk menyederhanakan data berdimensi tinggi menjadi sejumlah komponen utama (*principal components*) tanpa menghilangkan sebagian besar informasi penting yang terkandung dalam data. PCA mentransformasikan variabel-variabel yang saling berkorelasi menjadi variabel baru yang tidak saling berkorelasi, sehingga mampu mempertahankan variasi data terbesar pada beberapa komponen utama[41].

Dalam proses *data mining*, PCA sering digunakan sebagai tahap *preprocessing* sebelum proses pemodelan untuk mengurangi kompleksitas data, menghilangkan redundansi informasi, serta meningkatkan efisiensi komputasi. Dengan jumlah fitur yang lebih sedikit namun tetap mempertahankan sebagian besar informasi, proses analisis dapat dilakukan dengan lebih efektif.

Pada penelitian ini, PCA diterapkan setelah proses *feature scaling* dengan parameter **n_components = 0.95**, sehingga komponen utama yang dihasilkan mampu mempertahankan sekitar **95% variasi data**. Hasil transformasi PCA kemudian digunakan sebagai masukan pada proses Elbow Method, Silhouette Score, dan K-Means Clustering.

2.3 Framework/Algoritma yang digunakan

2.3.1 Framework CRISP-DM

Cross Industry Standard Process for Data Mining (CRISP-DM) merupakan salah satu framework yang paling banyak digunakan dalam penelitian data mining karena menyediakan tahapan kerja yang sistematis dan terstruktur dalam proses analisis data. Framework ini pertama kali dikembangkan oleh konsorsium perusahaan yang bergerak di bidang data mining untuk menyediakan standar metodologi yang dapat diterapkan pada berbagai bidang penelitian dan industri. CRISP-DM membantu peneliti dalam mengelola proses analisis data secara terarah, mulai dari memahami permasalahan penelitian hingga menghasilkan informasi yang dapat digunakan sebagai dasar pengambilan keputusan [13].

Menurut Saputra dan Yuliana [21], CRISP-DM terdiri atas enam tahapan utama yaitu Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, dan Deployment. Keenam tahapan tersebut saling terhubung dan memungkinkan peneliti untuk kembali ke tahap sebelumnya apabila ditemukan permasalahan selama proses analisis. Pendekatan ini memberikan fleksibilitas dalam pengolahan data sekaligus memastikan bahwa hasil yang diperoleh tetap sesuai dengan tujuan penelitian.

Pada penelitian ini, framework CRISP-DM digunakan sebagai pedoman dalam menganalisis persebaran minat dan karakteristik masyarakat terhadap olahraga padel di wilayah Jabodetabek. Tahapan Business Understanding digunakan untuk mengidentifikasi tujuan penelitian dan kebutuhan analisis. Tahapan Data Understanding dilakukan untuk memahami karakteristik data hasil survei. Selanjutnya, Data Preparation digunakan untuk membersihkan, mentransformasi, dan mempersiapkan data sebelum dilakukan proses pemodelan. Tahap Modeling dilakukan dengan menerapkan algoritma K-Means Clustering untuk mengelompokkan responden berdasarkan karakteristik yang dimiliki.

Hasil pengelompokan kemudian dievaluasi pada tahap Evaluation sebelum akhirnya disajikan dalam bentuk visualisasi dan interpretasi hasil pada tahap Deployment [37].

2.3.2 Algoritma K-Means Clustering

Pada penelitian ini, metode K-Means Clustering diterapkan untuk mengelompokkan responden yang memiliki karakteristik serupa berdasarkan tingkat minat, frekuensi bermain, serta berbagai aktivitas yang berkaitan dengan olahraga padel di wilayah Jabodetabek [14]. Melalui proses pengelompokan tersebut, responden dapat dibagi ke dalam beberapa kelompok yang merepresentasikan pola perilaku dan tingkat keterlibatan yang berbeda terhadap olahraga padel.

Tahapan clustering diawali dengan penentuan jumlah cluster yang akan dibentuk. Setelah jumlah cluster ditetapkan, algoritma akan menentukan centroid awal dan menghitung kedekatan setiap data terhadap centroid menggunakan metode Euclidean Distance [23]. Proses ini dilakukan secara berulang hingga diperoleh hasil pengelompokan yang stabil, di mana setiap responden berada pada cluster yang memiliki karakteristik paling sesuai dengan data yang dimilikinya.

Hasil clustering diharapkan mampu memberikan gambaran mengenai segmentasi masyarakat berdasarkan tingkat ketertarikan terhadap olahraga padel. Selain itu, metode ini juga dapat membantu mengidentifikasi pola-pola tertentu yang muncul pada setiap kelompok responden, sehingga analisis mengenai persebaran minat masyarakat terhadap olahraga padel dapat dilakukan secara lebih mendalam dan terarah.

2.4 Tools/software yang digunakan

2.4.1 Google Form

Google Form digunakan sebagai alat utama dalam proses pengumpulan data penelitian. Google Form memungkinkan peneliti untuk menyusun dan menyebarkan kuesioner secara digital serta mengumpulkan data responden secara efisien dan terstruktur [27].

Selain itu, Google Form memudahkan proses rekapitulasi data karena hasil jawaban responden dapat diekspor ke dalam format spreadsheet atau CSV yang mendukung proses analisis data lebih lanjut [26].

2.4.2 Microsoft Excel

Microsoft Excel atau Google Sheets digunakan sebagai alat bantu awal dalam proses pengecekan dan pengelolaan data hasil kuesioner. Tools ini digunakan untuk melakukan pemeriksaan awal terhadap data, seperti identifikasi data kosong, duplikasi data, serta penyesuaian format sebelum dilakukan analisis lanjutan menggunakan Python [21].

Penggunaan spreadsheet membantu peneliti dalam memahami struktur data secara awal dan memastikan data berada dalam kondisi yang siap untuk diproses.

2.4.3 Bahasa Pemrograman Python

Python digunakan sebagai tools utama dalam proses pengolahan dan analisis data pada penelitian ini. Python banyak digunakan dalam bidang Big Data Analytics karena memiliki sintaks yang sederhana, fleksibel, serta didukung oleh berbagai pustaka analisis data yang kuat [28].

Dalam penelitian ini, Python dimanfaatkan untuk melakukan analisis deskriptif, analisis sentimen, clustering, classification, Association Rule Mining, serta pengolahan data kuesioner secara terstruktur [12].

Python juga digunakan dalam implementasi machine learning dan data mining karena mendukung berbagai metode analisis data seperti K-Means Clustering, Decision Tree Classification, dan algoritma Apriori [14][15][16].

2.4.4 Library Python Pendukung

Beberapa library Python digunakan untuk mendukung proses analisis data dan visualisasi, antara lain pandas, NumPy, matplotlib, scikit-learn, dan mlxtend [29].

Library pandas digunakan untuk pengolahan dan manipulasi data, NumPy digunakan untuk mendukung proses komputasi numerik, sedangkan matplotlib digunakan untuk menyajikan data dalam bentuk grafik dan diagram [29].

Selain itu, library scikit-learn digunakan dalam implementasi metode K-Means Clustering dan Classification, sedangkan library mlxtend digunakan dalam proses Association Rule Mining menggunakan algoritma Apriori [25].

Penggunaan library Python tersebut membantu meningkatkan efisiensi dan akurasi proses analisis data.

2.4.5 Power BI

Power BI digunakan sebagai tools visualisasi data untuk menyajikan hasil analisis penelitian dalam bentuk dashboard interaktif, grafik, dan diagram [11]. Power BI membantu proses interpretasi data sehingga hasil analisis dapat dipahami secara lebih mudah dan informatif.

Dalam penelitian ini, Power BI digunakan untuk memvisualisasikan hasil analisis deskriptif, analisis sentimen, clustering, classification, serta Association Rule Mining terkait minat masyarakat terhadap olahraga padel di wilayah Jabodetabek [20].

Penggunaan Power BI memungkinkan peneliti menyajikan insight penelitian secara lebih komunikatif melalui tampilan dashboard yang interaktif dan mudah dipahami oleh pengguna.