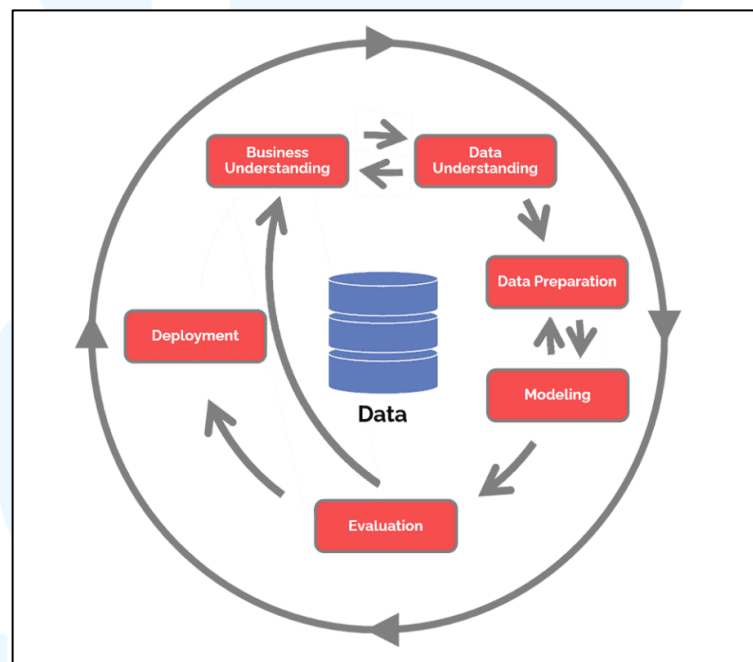


BAB III

METODE PENELITIAN

3.1 Metode Penelitian

Metode pengembangan sistem yang digunakan dalam penelitian ini adalah CRISP-DM. CRISP-DM merupakan metode *data mining* yang berfungsi untuk melakukan analisis data secara bertahap mulai dari mengerti masalah yang ingin diselesaikan hingga menangani solusi dari masalah tersebut dengan memanfaatkan data [32], [33]. Metode ini terdiri dari 6 tahap, yaitu *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modeling*, *Evaluation*, dan *Deployment*. Gambar di bawah ini menggambarkan metode CRISP-DM dalam bentuk diagram.



Gambar 3.1 Metode CRISP-DM [32]

Metode ini dipilih karena memiliki metode yang sejalan dengan perancangan *chatbot* untuk situs web Desa Pete, Tigaraksa. Berikut adalah perbandingan metode CRISP-DM dengan metode lainnya yang memiliki kesamaan.

Tabel 3.1 Tabel Perbandingan Metode

Aspek	CRISP-DM [34]	SEMMA [35]	KDD [35]
-------	---------------	------------	----------

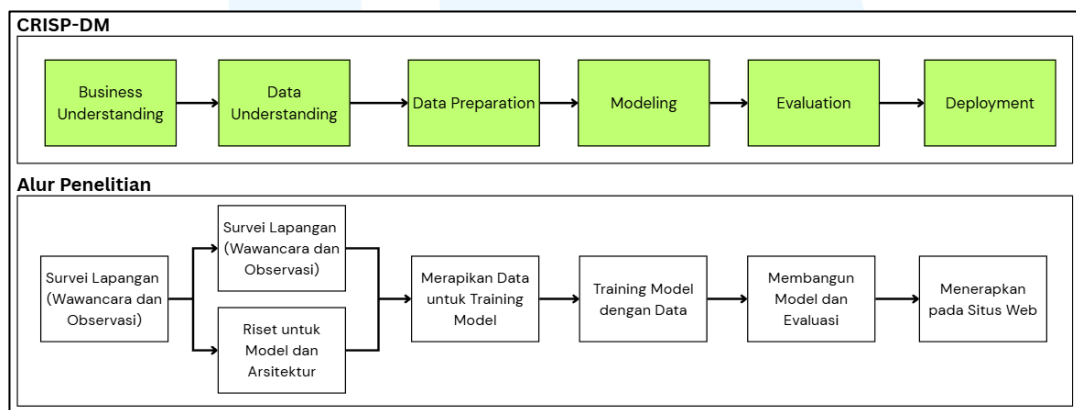
Deskripsi	Metodologi lintas industri yang dirancang untuk menghubungkan tujuan bisnis dengan proses analisis data secara sistematis dan iteratif.	Metodologi yang dikembangkan oleh SAS dan berfokus pada proses teknis pemodelan data secara eksploratif.	Kerangka kerja yang menekankan proses penemuan pengetahuan baru dari basis data besar melalui tahapan bertahap.
Metode	<i>Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, Deployment</i>	<i>Sample, Explore, Modify, Model, Assess</i>	<i>Business Understanding, Data Acquisition & Understanding, Modeling, Deployment, Customer Acceptance</i>
Kelebihan	1) Mengintegrasikan aspek bisnis dan teknis secara seimbang [36] 2) Mendukung proses <i>deployment</i> secara eksplisit	1) Alur teknis sederhana dan terstruktur 2) Efisien untuk eksperimen model dan eksplorasi data	1) Fokus kuat pada proses penemuan pengetahuan 2) Cocok untuk penelitian eksploratif dan akademik
Kekurangan	1) Membutuhkan pemahaman konteks bisnis yang matang 2) Implementasi bisa lebih kompleks dan memakan waktu	1) Tidak mencakup tahap <i>deployment</i> secara eksplisit 2) Kurang memperhatikan tujuan bisnis	1) Kurang terstruktur untuk implementasi sistem nyata 2) Tidak selalu mendefinisikan siklus iterasi dengan jelas
Kecocokan	Sangat sesuai untuk pengembangan sistem berbasis AI dan <i>chatbot</i> yang digunakan secara langsung oleh pengguna, termasuk sistem informasi pariwisata dan layanan publik [37].	Lebih cocok untuk analisis data internal, eksperimen model, dan studi teknis tanpa kebutuhan implementasi produksi.	Tepat digunakan dalam penelitian akademik atau eksplorasi data untuk menghasilkan wawasan dan teori baru.

Hasil peninjauan menunjukkan bahwa CRISP-DM dinilai paling sesuai untuk penelitian ini karena mampu memenuhi kebutuhan teknis dan tujuan bisnis dari awal hingga akhir proses. Metodologi ini mendukung proses *deployment* dan evaluasi berkelanjutan, yang sangat penting dalam pengembangan *chatbot* pariwisata Desa Pete. Hal ini sejalan dengan temuan penelitian sebelumnya yang menyatakan bahwa CRISP-DM lebih adaptif untuk sistem berbasis layanan publik dan aplikasi dunia nyata dibandingkan SEMMA dan KDD. SEMMA lebih menekankan aspek eksplorasi dan pemodelan data, sehingga kurang optimal untuk sistem yang membutuhkan integrasi langsung dengan pengguna akhir. Sementara itu, KDD lebih cocok digunakan sebagai kerangka konseptual dalam penelitian

eksploratif, namun tidak secara eksplisit mengatur implementasi sistem dalam lingkungan operasional.

3.2 Tahapan Penelitian

Penelitian yang dilakukan di Desa Pete, Tigaraksa melalui beberapa tahap untuk membangun *chatbot* di situs web Kecamatan Tigaraksa. Penjelasan alur penelitian dapat dilihat dalam *flowchart* berikut ini mengikuti metode CRISP-DM.



Gambar 3.2 Flowchart Alur Penelitian sesuai CRISP-DM

Penelitian dimulai dari perencanaan untuk melaksanakan penelitian berupa membuat lini masa setiap tahapan yang akan dilakukan, baik dari tahap persiapan penelitian hingga evaluasi penggunaan *chatbot*. Secara lebih spesifik, lini masa digambarkan pada Tabel 3.2 dalam bentuk *Gantt Chart*.

Tabel 3.2 Lini Masa Pelaksanaan Penelitian

No .	Deskripsi Pekerjaan	Agustus		September				Oktober				November				Desember			
		3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
1	Riset mengenai <i>chatbot</i> dan pelatihan yang digunakan terhadap model																		
2	Melakukan survei ke Desa Pete, Tigaraksa																		

No .	Deskripsi Pekerjaan	Agustus		September				Oktober				November				Desember			
		3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
3	Melakukan perencanaan pembangunan <i>chatbot</i> beserta fitur-fiturnya																		
4	Mengaplikasikan <i>chatbot</i> pada situs web Kecamatan Tigaraksa																		
5	Perancangan tabel pada <i>database</i> dan menguji coba <i>chatbot</i> pada situs web Kecamatan Tigaraksa																		
6	Penutupan proyek dan presentasi ke pihak desa																		

Pelaksanaan penelitian perancangan *chatbot* dimulai dari riset mengenai *chatbot* secara spesifik di bidang pariwisata beserta model dan pelatihan yang digunakan. Tahap ini dimulai pada minggu ke-3 Agustus. Kemudian dilanjutkan dengan survei ke Desa Pete, Tigaraksa pada minggu ke-4 Agustus untuk mengetahui keadaan dan kebutuhan dari pihak Desa. Berikut adalah dokumentasi saat melakukan survei ke Desa Pete, Tigaraksa.

UNIVERSITAS
MULTIMEDIA
NUSANTARA



Gambar 3.3 Diskusi dengan Pihak Desa Pete



Gambar 3.4 Survei Keliling Lingkungan Desa Pete



Gambar 3.5 Tempat Diskusi Penelitian

Penelitian dilanjutkan lagi ke tahap selanjutnya dari minggu ke-1 September hingga minggu ke-3 Oktober dengan melakukan perencanaan pembangunan *chatbot*. Perencanaan ini meliputi pengumpulan informasi mengenai desa dan riset terhadap model-model yang cocok. Riset ini dapat dilihat pada bab 2 penelitian di mana artikel-artikel mengusulkan banyak model dan arsitektur untuk *chatbot*. Selain itu, pada tahap ini dilakukan juga uji coba pelatihan model dan mengaplikasikannya dalam bentuk *chatbot*. Tahap ini mengambil waktu paling banyak dikarenakan banyak tantangan dan limitasi yang dihadapi selama pelaksanaannya.

Penelitian tahap selanjutnya dimulai dari minggu ke-4 Oktober hingga minggu ke-1 November. Pada tahap ini, pengaplikasian *chatbot* dalam bentuk API dilakukan terhadap situs web Kecamatan Tigaraksa yang sudah ada dari penelitian sebelumnya. Kemudian dilanjutkan dengan perancangan tabel pada *database* situs web Kecamatan Tigaraksa agar dapat diaplikasikan terhadap *chatbot* pada minggu ke-2 November hingga minggu ke-3 Desember. Tahap ini mengambil waktu yang cukup lama karena pengaplikasian *chatbot* cukup sulit saat memanggil API dan menghubungkan dengan basis data. Terakhir, penutupan proyek perancangan *chatbot* dilakukan dengan *deployment* pada situs web Kecamatan Tigaraksa serta

penulisan dan pengumpulan laporan dan artikel jurnal pada minggu ke-3 dan ke-4 Desember.

Pada nomor 3, 4, 5, dan 6, tahapan perancangan akan dijelaskan lebih terperinci menggunakan metode CRISP-DM. Berikut adalah penjelasannya.

3.2.1 *Business Understanding*

Pada tahap ini, identifikasi masalah dilakukan dengan tujuan untuk mengerti permasalahan yang perlu diselesaikan dan tujuan untuk menyelesaikan masalah tersebut [32]. Pada penelitian ini, identifikasi permasalahan utamanya adalah akses informasi pariwisata Desa Pete tersebut kepada pengunjung belakang masih terbatas, serta informasi digital berupa informasi pariwisata ini masih belum tersedia. Oleh karena itu, tujuan utama dari penelitian ini adalah mengembangkan sistem pariwisata berbasis AI untuk memberikan informasi tepat dan akurat serta memperluas paparan wisata nusantara dan pengetahuan teknologi.

3.2.2 *Data Understanding*

Pada tahap ini, diidentifikasi data yang diperlukan untuk menyelesaikan masalah yang sudah diidentifikasi pada tahap sebelumnya [32]. Pada penelitian ini, data diperoleh dari beberapa sumber utama, yaitu hasil survei langsung di Desa Pete, informasi yang dipublikasikan melalui berita daring, serta konten yang tersedia pada situs web resmi Kecamatan Tigaraksa yang memuat informasi terkait Desa Pete. Data hasil survei mencakup gambaran umum desa, fasilitas pariwisata, serta aktivitas yang ditawarkan kepada pengunjung. Sementara itu, data dari sumber berita dan situs web kecamatan digunakan sebagai data pendukung untuk melengkapi dan memverifikasi informasi desa yang telah dikumpulkan. Seluruh data tersebut kemudian dihimpun dan dianalisis untuk memastikan kesesuaian, kelengkapan, serta konsistensi informasi sebelum dimasukkan ke dalam database. Pada tahap ini juga dilakukan peninjauan awal terhadap potensi duplikasi, ketidaksesuaian konteks, dan kekosongan data yang dapat memengaruhi akurasi respons

chatbot. Proses pemahaman data ini menjadi dasar penting dalam memastikan bahwa informasi yang disajikan oleh sistem bersifat faktual, relevan, dan sesuai dengan kondisi nyata Desa Pete.

Selain dari sisi mengidentifikasi data yang diperlukan untuk analisis, peninjauan literatur juga dilakukan untuk mengetahui metode, arsitektur, dan model yang kompatibel untuk digunakan. Peninjauan literatur menunjukkan bahwa kombinasi arsitektur *Retrieval-Augmented Generation* (RAG) dan model bahasa yang efisien merupakan pendekatan yang paling menjanjikan untuk sistem *chatbot* dalam domain pariwisata desa, yang mengutamakan keakuratan konten faktual sekaligus konsistensi konteks. Arsitektur RAG menggabungkan mekanisme *retrieval* informasi dari knowledge base dengan kemampuan generation model bahasa, sehingga setiap respons disandarkan pada dokumen yang relevan dan tervalidasi, menekan risiko halusinasi yang seringkali terjadi pada sistem *generative-only*. Studi juga menunjukkan adanya evolusi tren dari sistem berbasis aturan (*rule-based*) menuju sistem RAG yang didukung embedding semantik, karena pendekatan ini mempertahankan relevansi konteks dan meningkatkan kepercayaan pengguna. Di sisi model bahasa, penelitian menunjukkan bahwa model berukuran menengah seperti GPT-2, terutama ketika dilakukan *fine-tuning* pada data domain spesifik, mampu memberikan respons koheren meskipun dengan kebutuhan komputasi yang lebih rendah dibandingkan large language models yang besar. Selain itu, model ini relatif mudah disesuaikan dengan variasi linguistik lokal dan memori operasional yang rendah, sehingga sesuai untuk lingkungan dengan keterbatasan sumber daya teknologi, seperti desa wisata yang ingin mengembangkan layanan informasi digital yang dapat diakses publik.

3.2.3 Data Preparation

Pada tahap ini, data diolah dan dipersiapkan untuk melakukan pemodelan [32]. Pada penelitian ini, data teks dibersihkan menjadi format yang sesuai agar dapat dibaca, dimengerti, dan dipelajari oleh arsitektur model bahasa. Data

percakapan disusun dalam format *prompt* dan *completion* untuk mendukung proses *fine-tuning*. Selain itu, data pengetahuan desa diproses menjadi representasi vektor menggunakan teknik *embedding* untuk mendukung mekanisme *retrieval* dalam arsitektur RAG. Tahap ini menjadi krusial karena kualitas data yang dipersiapkan sangat memengaruhi akurasi jawaban oleh sistem secara keseluruhan.

3.2.4 Modeling

Pada tahapan ini, data yang telah diolah mengalami proses pemodelan [32]. Tahapan ini melaksanakan implementasi model dan arsitektur yang telah ditentukan dalam fase pemahaman dari hasil tinjauan literatur. Skenario sistem terkait *chatbot* proses alur sistem, pemrosesan masukan user, serta sistem penyerahan jawaban dirancang untuk melahirkan respons yang informatif serta konsisten. Data yang telah melalui proses tahapan data preparation diubah ke dalam format JSON untuk dapat diproses secara efektif dalam model bahasa dalam proses pelatihan serta inferensi, selain itu juga untuk melindungi integrasi dari basis pengetahuan (*knowledge_base*) yang dimiliki sistemnya.

Pada model bahasa di tahap ini, model tersebut memiliki peran untuk mengembangkan jawaban berdasarkan data yang telah dilatih tersebut dan informasi yang dimiliki dalam *knowledge base*. Sistem tersebut telah mengoptimalkan arsitektur RAG untuk memberikan kepastian setiap jawaban tidak hanya dipengaruhi oleh daya generatif model tersebut, akan tetapi juga berdasarkan dokumen faktual dari hasil pengamatan dan penginterviewan yang telah diverifikasi. Model ini seiring dengan hasil penelitian masa lalu bahwa kombinasi sistem arsitektur berbasis dokumen dan model bahasa berbasis *generation* berkontribusi nyata terhadap ketepatan jawaban sistem berbasis *chatbot* informasi publik.

Lebih lanjut, dari literatur yang lebih baru, terlihat bahwa implementasi dari model yang berukuran kecil sampai menengah yang dikombinasikan dengan struktur arsitektur RAG dapat memberikan performa yang

berkompetisi tanpa perlu memproses sumber daya komputasi yang besar. Karena itu, dalam tahapan pemodelan, model ini hanya digunakan sebagai mesin penjawab yang berfungsi demikian dengan cara memanfaatkan data dari hasil pelatihan pengetahuan dan penelusuran untuk memberitahu berita tentang kegiatan pariwisata desa secara tepat, relevan, serta aman dari kemungkinan halusinasi. Karena itu, relasi antara tahapan pemodelan berpengaruh sebagai

3.2.5 Evaluation

Tahap ini berfungsi untuk menilai seberapa tinggi akurasi dan keberhasilan sistem yang sudah dirancang untuk memenuhi tujuan pada tahap *business understanding* [32]. Evaluasi dilakukan menggunakan beberapa metrik, antara lain *Retrieval Success Rate*, tingkat penggunaan model, *Hallucination Rate*, serta metrik kesamaan semantik seperti *BERTScore*. Selain evaluasi kuantitatif, dilakukan juga pengujian pada situs web Kecamatan Tigaraksa di halaman Desa Pete, yaitu pengumpulan data pengguna, kemampuan menjawab berbasis *database*, serta pencatatan pertanyaan yang tidak dapat dijawab.

3.2.6 Deployment

Tahap ini merupakan tahap terakhir di mana *chatbot* yang telah dirancang dan diuji diintegrasikan ke dalam situs web Kecamatan Tigaraksa pada halaman Desa Pete [32]. Sistem dijalankan melalui backend berbasis FastAPI yang menghubungkan pengguna dengan *chatbot*. Pada tahap ini, *chatbot* dapat diakses langsung oleh pengunjung situs dan berinteraksi secara *real-time*. Selain itu, sistem juga mencatat pertanyaan yang tidak dapat dijawab ke dalam *database* untuk mendukung menyimpan informasi untuk dikembangkan lebih lanjut.

3.3 Teknik Pengumpulan Data

Data yang digunakan dalam penelitian ini merupakan kombinasi antara data primer dan data sekunder untuk memastikan kelengkapan serta keakuratan informasi yang mendukung pengembangan sistem *chatbot*. Data primer diperoleh

melalui kegiatan survei lapangan yang dilakukan secara langsung di Desa Pete. Survei ini bertujuan untuk mengumpulkan informasi mengenai profil atau informasi umum desa, potensi pariwisata, fasilitas pendukung, serta kebutuhan informasi yang paling sering ditanyakan oleh pengunjung maupun masyarakat setempat. Data primer tersebut dikumpulkan untuk menjadi sumber informasi yang akan digunakan pada *database*.

Data sekunder juga digunakan untuk mengumpulkan informasi tambahan mengenai Desa. Data dikumpulkan dari berbagai sumber pendukung yang telah tersedia sebelumnya, seperti informasi yang dipublikasikan melalui media berita daring serta konten resmi yang terdapat pada situs web Kecamatan Tigaraksa yang memiliki informasi terkait Desa Pete. Data sekunder ini dimanfaatkan untuk memperbanyak data primer sehingga informasi yang digunakan dalam sistem *database*. Penggabungan data primer dan data sekunder ini dilakukan untuk memperbanyak informasi, memperbanyak pengetahuan untuk dipelajari *chatbot*, serta memastikan bahwa jawaban yang dihasilkan oleh *chatbot* fakta dan relevan.

3.4 Teknik Analisis Data

Teknik analisis data yang digunakan dalam penelitian ini adalah pendekatan kualitatif dengan metode observasi lapangan dan wawancara. Pendekatan ini dipilih dan dilakukan pada saat melakukan survei di awal pelaksanaan penelitian yang ditunjukkan pada tabel lini masa di nomor 2, yaitu melakukan survei ke Desa Pete, Tigaraksa.

Observasi lapangan dilakukan secara langsung di Desa Pete untuk mengamati wisata yang dapat dikunjungi, fasilitas yang tersedia, aktivitas masyarakat, dan interaksi antara pengelola desa dan pengunjung. Sementara itu, wawancara dilakukan dengan pihak desa sebagai informan utama untuk mengumpulkan informasi mengenai profil desa, letak tempat wisata beserta deskripsinya, dan fasilitas yang disediakan. Hasil observasi dan wawancara tersebut dianalisis secara deskriptif untuk melatih model yang akan dipakai agar dapat mempelajari informasi.