

BAB III

METODOLOGI PENELITIAN

3.1 Gambaran Umum Objek Penelitian

Objek penelitian ini adalah proses generasi dan evaluasi gambar edukatif berbasis teknologi *Text-to-Image (T2I) Generation* dalam konteks pembelajaran anak-anak di Desa Wisata Tigaraksa, Kabupaten Tangerang. Penelitian ini berfokus pada bagaimana kecerdasan generatif dapat dimanfaatkan untuk menciptakan media pembelajaran visual yang kontekstual, menarik dan mudah dipahami oleh siswa sekolah dasar tanpa memerlukan keterampilan desain profesional.

Objek utama penelitian mencakup hasil visual yang dihasilkan oleh model T2I berdasarkan variasi metode *prompt engineering* yang disusun secara sistematis. Dalam penelitian ini, setiap gambar yang dihasilkan diperlakukan sebagai satu *instance* untuk evaluasi, di mana variasi prompt (descriptive, instruction-based, dan compositional) akan dibandingkan untuk menilai efektivitasnya dalam menghasilkan visual edukatif yang sesuai dengan konteks lokal.

Tema yang digunakan dalam penelitian ditentukan berdasarkan potensi dan kegiatan edukasi di Desa Tigaraksa, antara lain meliputi topik flora dan fauna lokal dan proses pertanian. Seluruh tema dirancang agar relevan dengan pembelajaran anak-anak sekolah dasar.

Penelitian ini menggunakan dua model T2I sebagai bandingan yaitu: Gemini Nano Banana, model multimodal berbasis bahasa alami yang unggul dalam pemahaman semantik dan dukungan terhadap bahasa Indonesia dan Stable Diffusion , model *open-source* berbasis *latent diffusion* yang banyak digunakan dalam riset visual generatif karena fleksibilitasnya dan kemampuannya menghasilkan detail artistik dengan kontrol yang tinggi.

Setiap kombinasi metode prompt dan model T2I akan menghasilkan serangkaian gambar yang dievaluasi melalui dua pendekatan: (1) evaluasi kuantitatif menggunakan metrik *CLIPScore* untuk mengukur kesesuaian semantik dan distribusi visual serta (2) evaluasi kualitatif berbasis persepsi manusia melalui pengelola desa dan juga pendidik.

Sebagai luaran akhir, penelitian ini akan merancang prototipe website yang memungkinkan pengguna (guru, siswa, maupun pengelola desa) untuk menghasilkan gambar edukatif secara mandiri hanya dengan menuliskan deskripsi teks. Dengan demikian, objek penelitian ini tidak hanya bersifat teknis tetapi juga edukatif dan sosial. Dari sisi teknis, penelitian ini menguji efektivitas metode *prompt engineering* dalam menghasilkan gambar yang relevan dengan teks deskriptif. Dari sisi edukatif dan sosial, penelitian ini berkontribusi dalam pengembangan media pembelajaran yang sesuai bagi anak-anak desa.

3.1.2 Jenis dan Pendekatan Penelitian

Penelitian ini menggunakan pendekatan kuantitatif-komparatif dengan mixed methods. Pendekatan kuantitatif diterapkan melalui pengukuran objektif menggunakan metrik *CLIPScore* untuk mengevaluasi kesesuaian antara prompt dan gambar yang dihasilkan oleh model *Text-to-Image (T2I)*. Sementara itu, pendekatan komparatif digunakan untuk membandingkan performa dua model T2I yaitu Gemini Nano Banana dan Stable Diffusion berdasarkan tiga variasi metode prompt engineering (descriptive, instructive, dan compositional).

Dan penelitian ini juga mengintegrasikan metode kualitatif melalui evaluasi berbasis persepsi manusia (human evaluation). Pendekatan mixed methods ini dipilih karena kualitas visual edukatif tidak dapat dinilai secara komprehensif hanya melalui metrik matematis tetapi juga memerlukan penilaian dari pengguna akhir. Dalam hal ini pendidik, pengelola desa wisata dan anak-anak sebagai audiens sasaran.

Jenis penelitian ini dapat dikategorikan sebagai *experimental comparative research*, di mana eksperimen dilakukan dengan mengendalikan variabel-variabel

tertentu (tema, parameter model, dan seed) untuk mengisolasi pengaruh metode prompt terhadap kualitas visual yang dihasilkan. Dengan demikian, penelitian ini tidak hanya menghasilkan deskripsi fenomena, tetapi juga mengidentifikasi hubungan antara variasi *prompt engineering* dan kualitas output visual dalam konteks pembelajaran berbasis komunitas.

3.1.3 Variabel Penelitian

Penelitian ini melibatkan tiga kategori variabel yang didefinisikan untuk menjaga validitas eksperimen. Variabel independen adalah metode prompt engineering yang terdiri dari tiga kategori yaitu *Descriptive Prompt* (penjelasan detail visual seperti warna, bentuk, dan suasana), *Instruction Prompt* (perintah eksplisit untuk membuat poster atau infografis edukatif) dan *Compositional Prompt* (pengaturan tata letak dan hubungan antar elemen visual). Masing-masing metode diuji dengan 10 variasi prompt. Ketiga metode ini dipilih berdasarkan kajian literatur yang menunjukkan bahwa variasi struktur prompt memengaruhi kualitas output T2I secara signifikan terutama dalam konteks pembelajaran visual untuk anak-anak.

Variabel dependen adalah kualitas visual edukatif yang diukur melalui empat dimensi. Pertama, Relevansi Kontekstual mengukur kesesuaian gambar dengan tema pembelajaran dan konteks lokal Desa Wisata Tigaraksa menggunakan CLIPScore dan *Human Evaluation* menggunakan penilaian panel (skala Likert 1-5). Kedua, Akurasi untuk mengukur ketepatan visual dalam merepresentasikan objek dan konsep yang dijelaskan dalam prompt. Ketiga, Estetika Visual mengukur kualitas komposisi, keseimbangan warna dan daya tarik visual gambar.

Variabel kontrol dijaga konstan untuk memastikan validitas internal. Tema Pembelajaran dibatasi pada tiga kategori yaitu Fauna Lokal, Flora Perkebunan, dan Aktivitas Pertanian. Parameter Model untuk Stable Diffusion dijaga konstan (Guidance Scale 7.5, Inference Steps 50, Sampler Euler), sementara Gemini Nano Banana bekerja secara otomatis tanpa pengaturan manual. *Seed Value* ditetapkan pada angka 42 untuk memastikan replikasi hasil. Resolusi Output dijaga pada 512×512 piksel sebagai standar untuk kedua model. Bahasa Prompt menggunakan Bahasa Indonesia secara konsisten untuk menguji kemampuan model dalam

memahami konteks lokal dan bahasa indonesia. Dengan mengontrol variabel-variabel ini, penelitian dapat mengisolasi pengaruh metode prompt terhadap kualitas visual tanpa bias dari faktor luar.

3.2 Metode Penelitian

3.2.2 Pemilihan Metode Team Data Science Processing

Penelitian ini mempertimbangkan dua kerangka kerja yang paling banyak digunakan dalam proyek berbasis data, yaitu CRISP-DM dan *Team Data Science Process (TDSP)*. CRISP-DM dikenal sebagai metodologi yang efektif untuk data mining dan analisis eksploratif, terutama ketika fokus penelitian terletak pada proses pembersihan data, identifikasi pola, serta pemodelan. Namun, penelitian ini tidak berorientasi pada data mining tradisional, melainkan berfokus pada eksperimen model *Generative AI*, yang melibatkan pengujian beberapa metode prompt engineering, produksi visual secara iteratif dan evaluasi performa model menggunakan metrik teknis dan penilaian manusia. Dengan kebutuhan tersebut, *TDSP* dipilih karena menawarkan kerangka yang sudah selaras dengan alur penelitian berbasis AI modern mulai dari akuisisi data, pengembangan model, evaluasi hingga integrasi ke dalam sistem yang dapat digunakan oleh pengguna akhir.

Oleh karena itu, setelah membandingkan kedua metode di tabel 3.1, penelitian ini memilih menggunakan *TDSP* sebagai metode karena kemampuannya dalam mendukung proses penelitian yang bersifat eksperimental. Pendekatan *TDSP* memungkinkan penelitian ini untuk mengelola seluruh rangkaian proses mulai dari perumusan masalah, pengumpulan data, eksperimen T2I hingga evaluasi secara lebih terukur untuk penelitian lanjutan.

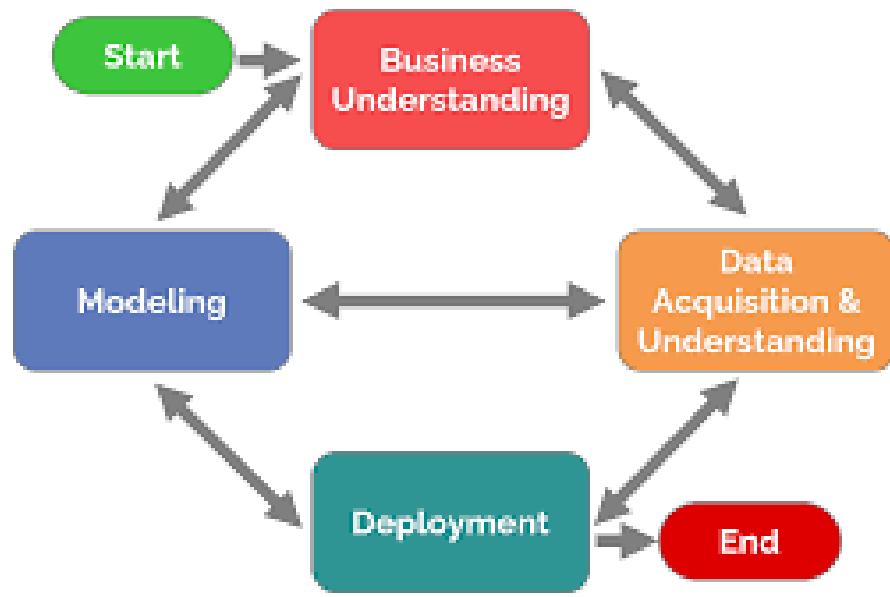
Tabel 3.1 Perbandingan Metode TDSP dan CRISP-DM

Aspek	TDSP	CRISP-DM
-------	------	----------

Tujuan	Fokus pada proyek AI/ML modern dengan pipeline end-to-end	Fokus pada data mining tradisional
Struktur	Lebih teknis, modular, mendukung iterasi dan eksperimen model	Lebih konseptual, berorientasi analisis
Tahapan	Business Understanding, Data Acquisition, Modeling, Deployment, Acceptance.	Business Understanding, Data Understanding, Preparation, Modeling, Evaluation, Deployment.
Kesesuaian dengan Penelitian	Sangat cocok untuk Generative AI, evaluasi model, dan eksperimen prompt.	Kurang relevan untuk T2I dan workflow AI modern.

Team Data Science Process (TDSP) merupakan kerangka kerja metodologis yang dikembangkan oleh Microsoft untuk membantu tim data sains menjalankan proyek secara sistematis, kolaboratif, dan terukur [26]. Pendekatan ini dirancang agar setiap tahapan penelitian mulai dari perencanaan hingga evaluasi dapat berjalan dengan struktur yang konsisten dan mudah direplikasi. Dalam konteks penelitian ini, TDSP digunakan sebagai panduan utama dalam mengelola proses eksperimen *Generative AI* berbasis *Text-to-Image (T2I)*, mulai dari pengumpulan data, pembangunan model hingga penerapan dan penilaian hasil.

UNIVERSITAS
MULTIMEDIA
NUSANTARA



Gambar 3.1 Team Data Science Process

Tahapan pertama dalam TDSP adalah *business understanding*, yaitu proses memahami tujuan dan kebutuhan proyek. Pada penelitian ini, tahap tersebut diterapkan untuk mengidentifikasi permasalahan Desa Wisata Tigaraksa dalam penyediaan media pembelajaran visual yang menarik dan relevan dengan konteks lokal. Tahap kedua adalah *data acquisition and understanding*, yang berfokus pada pengumpulan dan eksplorasi data berupa *prompt text* dan hasil gambar T2I. Seluruh data disimpan di Supabase agar dapat diatur, dipanggil dan dievaluasi secara efisien.

Tahap berikutnya adalah *modeling*, yaitu proses eksperimen menggunakan berbagai metode *prompt engineering* (descriptive, instruction-based, dan compositional) untuk menghasilkan gambar edukatif. Hasil dari tahap ini dievaluasi secara kuantitatif menggunakan *CLIPScore* serta secara kualitatif melalui penilaian persepsi manusia terhadap kualitas dan kesesuaian gambar. Setelah itu, tahap *deployment* dilakukan untuk mengintegrasikan seluruh hasil ke dalam Supabase, sehingga evaluasi dapat diakses secara real-time dan digunakan oleh pengguna non-teknis seperti guru atau pengelola desa.

Tahap terakhir adalah evaluasi pengguna. Pada penelitian ini, masyarakat dan pengelola Desa Wisata Tigaraksa menilai sejauh mana visual yang dihasilkan relevan, menarik serta bermanfaat dalam konteks pembelajaran dan promosi wisata edukatif. Melalui lima tahap tersebut, TDSP membantu memastikan bahwa setiap proses penelitian berjalan secara iteratif dan terukur. Hasil yang diperoleh tidak hanya bersifat eksperimental, tetapi juga memiliki nilai praktis dan dapat diterapkan kembali di konteks serupa. Dengan menerapkan TDSP, penelitian ini berupaya menghubungkan pendekatan teknis AI dengan kebutuhan nyata masyarakat, sekaligus membangun proses penelitian yang transparan, efisien dan berkelanjutan.

3.2.3 Alur Penelitian

Alur penelitian ini diawali dengan tahapan *business understanding*, yang bertujuan untuk memahami permasalahan inti dan kebutuhan nyata masyarakat Desa Wisata Tigaraksa. Pada tahap ini, dilakukan studi literatur mengenai penerapan *Generative Artificial Intelligence (GenAI)* dalam bidang pendidikan serta analisis *research gap* terkait penggunaan *Text-to-Image (T2I)* sebagai alat bantu pembelajaran visual.

Selain melalui analisis, tahap ini juga mencakup kegiatan observasi lapangan dan wawancara langsung bersama pengelola Desa Wisata Tigaraksa. Kunjungan lapangan dilakukan bersama tim penelitian untuk memperoleh pemahaman mengenai bagaimana kegiatan edukasi saat ini dilaksanakan. Temuan dari tahap ini menjadi dasar dalam perumusan fokus penelitian, yaitu mengembangkan sistem berbasis *Text-to-Image Generation* yang dapat membantu menciptakan gambar edukatif sesuai kebutuhan desa. Selain itu, hasil observasi juga digunakan untuk menyusun kriteria evaluasi visual, sehingga gambar yang dihasilkan tidak hanya dinilai secara teknis tetapi juga dari segi relevansi konteks dan daya tarik bagi anak-anak desa.

Tahap berikutnya adalah *data acquisition & understanding*, yang mencakup proses penentuan tema pembelajaran dan penyusunan dataset *prompt*. Tema-tema diambil dari aktivitas dan lingkungan Desa Tigaraksa seperti tanaman perkebunan

fauna lokal dan praktik pertanian yang ingin diajarkan. Setiap tema kemudian dikembangkan menjadi beberapa variasi *prompt* berdasarkan tiga metode utama, yaitu *descriptive prompt*, *instruction-based prompt*, dan *compositional prompt*. Proses ini dilakukan untuk memahami bagaimana variasi instruksi teks dapat memengaruhi kualitas gambar yang dihasilkan oleh model T2I.

Tahap selanjutnya adalah modeling yang berfokus pada proses *image generation* menggunakan dua model T2I, yaitu Gemini Nano Banana dan Stable Diffusion. Setiap kombinasi model dan metode *prompt* diuji dengan parameter yang dikendalikan untuk menjaga konsistensi hasil. Proses ini dilaksanakan menggunakan *Hugging Face Diffusers* dan *Jupyter Notebook* sebagai lingkungan eksperimen utama. Hasil gambar yang dihasilkan kemudian disimpan dalam repositori terstruktur sebagai dataset hasil generasi yang siap dievaluasi.

Setelah proses generasi selesai, tahap berikutnya adalah evaluation, yang dilakukan melalui dua pendekatan yaitu *objective evaluation* dan *subjective evaluation*. Evaluasi objektif mencakup penghitungan metrik kuantitatif seperti *CLIPScore* untuk menilai kesamaan distribusi visual dan kesesuaian semantik antara teks (*Prompt*) dan gambar. Sementara itu, evaluasi subjektif dilakukan melalui oleh pendidik dan pengelola Desa Tigaraksa untuk menilai aspek relevansi, daya tarik visual dan kesesuaian gambar dengan konteks pembelajaran anak. Hasil dari kedua pendekatan ini dibandingkan untuk menentukan kombinasi model dan metode *prompt* terbaik.

Tahap terakhir adalah *deployment*, di mana hasil penelitian diintegrasikan ke dalam prototype website. Website ini dirancang agar pengguna seperti guru atau pengelola desa dapat menuliskan deskripsi teks (*Prompt*) sederhana dan secara otomatis menghasilkan ilustrasi edukatif yang sesuai. Prototipe website ini sekaligus menjadi *proof of concept* dari penerapan hasil penelitian ke dalam sistem pembelajaran berbasis teknologi yang mudah digunakan dan dapat diakses masyarakat lokal.

Dengan mengikuti alur penelitian ini, diharapkan sistem yang dikembangkan tidak hanya menghasilkan temuan akademis mengenai efektivitas metode *prompt engineering* dalam konteks *Text-to-Image*, tetapi juga memberikan kontribusi berupa platform edukatif berbasis AI yang dapat digunakan secara langsung di Desa Wisata Tigaraksa. Pendekatan ini menjadikan penelitian bersifat aplikatif, berorientasi pada dampak sosial dan sekaligus memperkuat integrasi antara inovasi teknologi dan pembelajaran di tingkat komunitas.

3.2.4 Pengembangan Metode Prompt Engineering

Metode *prompt engineering* dalam penelitian ini dikembangkan untuk mengetahui pengaruh variasi struktur prompt terhadap kualitas hasil visual yang dihasilkan oleh model *Text-to-Image (T2I)* dalam konteks edukasi anak-anak. Berdasarkan hasil observasi di Desa Wisata Tigaraksa, penelitian ini menggunakan tiga pendekatan utama, yaitu *descriptive prompt*, *instruction-based prompt* dan *compositional prompt*.

Pendekatan *descriptive* berfokus pada penjabaran detail objek dan suasana, seperti warna, bentuk dan latar visual agar AI mampu menghasilkan gambar yang konkret dan mudah dipahami anak-anak. Pendekatan *instruction-based* digunakan untuk memberikan arahan eksplisit terhadap tujuan gambar, misalnya pembuatan infografis atau ilustrasi pembelajaran tertentu. Sedangkan *compositional prompt* berfungsi mengatur tata letak dan hubungan antar elemen visual agar hasil lebih seimbang dan komunikatif secara desain [17]

Ketiga pendekatan ini dipilih karena saling melengkapi: *descriptive prompt* memperkuat ketepatan visual, *instruction-based prompt* memastikan konsistensi makna dan fungsi edukatif sementara *compositional prompt* menjaga visual agar pesan pembelajaran tersampaikan secara efektif.

3.2.5 Pemilihan Model Text-to-Image

Pemilihan model *Text-to-Image (T2I)* didasarkan pada kebutuhan untuk menilai perbedaan pendekatan arsitektur AI dalam menghasilkan visual edukatif yang

relevan dengan konteks lokal. Dua model digunakan dalam penelitian ini, yaitu *Stable Diffusion* dan *Gemini Nano Banana (Flash)*.

Tabel 3.2 Perbandingan Model Text-to-Image

Aspek	Stable Diffusion 2.1	Gemini Nano Banana (Flash)
Arsitektur	Latent Diffusion Model (LDM)	Multimodal Transformer
Kelebihan	Open-source Kontrol parameter tinggi (guidance scale, steps, seed) Fleksibel untuk eksperimen Dapat dijalankan lokal	Pemahaman bahasa natural superior Dukungan bahasa Indonesia native Mudah digunakan non-teknis Inference cepat
Kelemahan	Kurang optimal untuk bahasa Indonesia Komputasi tinggi	Closed-source Tidak bisa kontrol parameter teknis
Platform Akses	<i>Hugging Face + Jupyter Notebook</i>	<i>Web Interface Gemini AI</i>
Ukuran Output	512×512 px (dapat disesuaikan)	512×512 px (otomatis)

Model Stable Diffusion 2.1 dipilih karena bersifat *open-source*, fleksibel, dan dapat dijalankan secara lokal, sehingga memberikan kontrol penuh terhadap parameter generasi seperti *guidance scale*, *inference steps*, dan *seed*. Model ini menggunakan arsitektur latent diffusion, di mana proses pembentukan gambar

dilakukan secara bertahap dari *latent space*, memungkinkan eksperimen yang sistematis, transparan, dan dapat direplikasi dalam konteks penelitian akademik. Selain itu, kemampuan untuk dijalankan tanpa ketergantungan pada API komersial menjadikannya relevan untuk lingkungan dengan keterbatasan infrastruktur.

Sebaliknya, Gemini Nano Banana dipilih karena mengadopsi pendekatan transformer-based multimodal model yang unggul dalam memahami konteks bahasa alami dan relasi teks–visual, termasuk deskripsi berbahasa Indonesia. Model ini lebih adaptif dalam menghasilkan visual yang sesuai dengan makna teks tanpa memerlukan *prompt engineering* teknis yang kompleks, sehingga lebih mudah digunakan oleh pengguna non-teknis seperti pendidik dan pengelola desa. Dengan demikian, Stable Diffusion merepresentasikan paradigma model yang menonjol dari sisi teknis dan fleksibilitas kustomisasi, sementara Gemini Nano Banana merepresentasikan paradigma model yang unggul dalam pemahaman semantik dan konteks bahasa, memungkinkan perbandingan dua pendekatan T2I yang berbeda namun saling melengkapi.

Model *text-to-image* lain seperti *DALL·E*, *Midjourney*, dan *Adobe Firefly* tidak dipilih terutama karena keterbatasannya dalam memahami prompt berbahasa Indonesia dan konteks lokal secara konsisten. Model-model tersebut umumnya menunjukkan performa yang lebih baik pada prompt berbahasa Inggris dan cenderung menghasilkan visual dengan interpretasi semantik yang kurang akurat ketika digunakan dalam konteks bahasa non-Inggris. Sebaliknya, Gemini Nano Banana dipilih karena kemampuannya dalam memproses bahasa alami Bahasa Indonesia secara lebih baik dan menghasilkan visual yang lebih selaras dengan makna prompt tanpa memerlukan *prompt engineering* teknis yang kompleks.

Pemilihan Gemini juga didukung oleh temuan penelitian Noel et al. [40] yang membandingkan empat model T2I, yaitu *DALL·E*, *Copilot*, *Gemini*, dan *Midjourney*, dalam konteks ilustrasi edukatif. Penelitian tersebut menunjukkan bahwa Gemini secara konsisten mengungguli model lain dalam aspek akurasi visual, konsistensi warna, dan nilai edukatif. Temuan ini sejalan dengan hasil

penelitian ini, di mana Gemini menunjukkan performa dalam menghasilkan visual edukatif berbasis prompt Bahasa Indonesia. Oleh karena itu, Gemini Nano Banana dipilih sebagai model pembanding utama karena relevansinya terhadap konteks bahasa, edukasi, dan temuan dari penelitian terdahulu.

3.2.6 Pemilihan Metriks Evaluasi Model T2I

Tahapan evaluasi dilakukan untuk menilai kesesuaian antara hasil gambar dan teks deskriptif dari prompt. Penelitian ini menggunakan dua jenis evaluasi, yaitu evaluasi objektif menggunakan metrik CLIPScore, dan evaluasi subjektif melalui human evaluation.

CLIPScore digunakan karena mampu menilai kesesuaian semantik antara teks dan gambar dengan menghitung *cosine similarity* antara vektor representasi keduanya dalam ruang embedding CLIP. Semakin tinggi nilai CLIPScore, semakin baik tingkat kesesuaian antara deskripsi teks dan hasil visual.

Sementara itu, human evaluation dilakukan oleh panel penilai yang terdiri dari guru sekolah dasar, pengelola desa wisata dan anak-anak. Mereka menilai empat aspek utama: (1) relevansi konteks lokal, (2) akurasi objek, (3) estetika visual, dan (4) kesesuaian dengan audiens anak-anak. Skala penilaian menggunakan skala Likert 1–5.

Kombinasi kedua pendekatan ini dipilih untuk memastikan hasil evaluasi tidak hanya mengukur performansi teknis model, tetapi juga kualitas edukatif dan persepsi manusia terhadap hasil visual.

3.2.7 Metode Data Mining

Metode penelitian yang digunakan dalam studi ini mengacu pada kerangka kerja *Team Data Science Process* (TDSP) yang dimodifikasi agar sesuai dengan konteks penelitian berbasis *Generative Artificial Intelligence* (GenAI). TDSP

dipilih karena menyediakan alur kerja yang sistematis, iteratif dan dapat diterapkan baik pada penelitian akademik maupun pengembangan sistem berbasis AI. Dalam konteks ini, TDSP tidak hanya digunakan terhadap model *Text-to-Image (T2I)*, tetapi juga untuk memastikan bahwa hasil penelitian memiliki dampak praktis terhadap kebutuhan masyarakat lokal di Desa Wisata Tigaraksa.

Kerangka TDSP ini terdiri dari lima tahap utama: *Business Understanding, Data Acquisition & Understanding, Modeling, Evaluation, dan Deployment*. Setiap tahapan dijalankan secara berulang dan saling terhubung, dengan evaluasi hasil di setiap tahap sebagai dasar penyempurnaan proses berikutnya.

3.2.7.1 Business Understanding

Tahapan ini bertujuan untuk memahami secara mendalam permasalahan inti, konteks sosial, serta kebutuhan nyata dari pengguna di lapangan. Peneliti bersama tim tidak hanya melakukan observasi umum, tetapi juga melaksanakan kunjungan langsung ke Desa Wisata Tigaraksa untuk melakukan *in-depth interview* dengan pengelola desa. Pendekatan ini memungkinkan peneliti memperoleh pemahaman tentang bagaimana nanti nya kegiatan edukatif dilaksanakan di lingkungan desa.

Selain itu, dilakukan identifikasi kebutuhan pengguna untuk mengetahui jenis visual yang dianggap relevan dan efektif dalam konteks pembelajaran anak. Kegiatan ini juga mencakup pengumpulan insight dari guru dan pengelola wisata edukasi terkait preferensi gaya visual, karakter lokal yang ingin ditonjolkan, serta potensi pemanfaatan gambar untuk kegiatan edukatif dan promosi. Hasil tahap ini menjadi dasar dalam merancang tema, struktur prompt dan tujuan teknis dari sistem T2I yang akan dikembangkan.

3.2.7.2 Data Acquisition & Understanding

Tahap ini berfokus pada perancangan dataset berupa kumpulan *prompt* dan gambar hasil generasi model T2I. Berdasarkan hasil observasi dan diskusi lapangan, ditentukan beberapa tema utama yang mencerminkan potensi lokal Desa Tigaraksa, seperti fauna lokal , flora perkebunan dan aktivitas pertanian.

Selanjutnya, setiap tema dikembangkan ke dalam tiga variasi metode prompt engineering, yaitu *descriptive*, *instruction-based* dan *compositional*. Masing-masing metode menghasilkan 10 variasi prompt, sehingga total terdapat 30 prompt yang akan diuji. Setiap prompt kemudian dijalankan pada dua model T2I yang berbeda yaitu *Stable Diffusion* (bersifat open-source, fleksibel, dan dapat dikontrol secara lokal) serta *Gemini Nano Banana* (model berbasis multimodal dengan dukungan bahasa alami yang kuat). Dengan demikian, total gambar yang dihasilkan mencapai 60 hasil generasi visual yang merepresentasikan kombinasi antara metode dan model yang diuji.

Tahap ini juga mencakup kegiatan preprocessing teks untuk memastikan struktur prompt konsisten (terdiri dari subject, context, style, dan tone) serta pengaturan parameter model seperti *guidance scale*, *inference steps*, dan *seed value* agar hasil antar eksperimen dapat dibandingkan secara objektif. Hasil dari tahap ini adalah kumpulan gambar edukatif yang siap digunakan untuk tahap modeling dan evaluasi.

Penelitian ini juga menggunakan dataset yang terdiri dari 30 *prompt* hasil perancangan dengan tiga metode *prompt engineering* (Descriptive, Instruction-Based, dan Compositional). Pemilihan 30 prompt dalam penelitian ini didasarkan pada pertimbangan metodologis untuk menjaga validitas, kedalaman analisis, dan kelayakan evaluasi. Dalam studi komparatif model generatif, jumlah sampel kecil–menengah (20–50 item) umum digunakan ketika fokus penelitian adalah analisis kualitas keluaran model, bukan pelatihan model skala besar. Dengan membagi 30 prompt ke dalam tiga kategori prompt engineering (Descriptive, Instruction-Based, dan Compositional), setiap kategori memperoleh representasi yang seimbang untuk dianalisis secara kuantitatif maupun kualitatif.

Selain itu, dua model menghasilkan total 60 gambar, yang secara metodologis dianggap ideal untuk penelitian yang melibatkan Human Evaluation. Jumlah ini menghindarkan panel penilai dari kelelahan kognitif yang dapat menurunkan reliabilitas penilaian apabila jumlah sampel terlalu besar. Dalam penelitian

berbasis persepsi visual, menjaga kualitas penilaian lebih penting daripada memperbanyak jumlah sampel secara tidak terkontrol.

3.2.7.3 Modelling

Tahapan modeling dilakukan untuk menganalisis dan membandingkan performa kedua model *Text-to-Image (T2I)* dalam menghasilkan gambar edukatif yang relevan dengan konteks lokal Desa Wisata Tigaraksa. Seluruh proses eksperimen dilakukan di lingkungan Jupyter Notebook menggunakan *library* Hugging Face Diffusers untuk menjalankan model Stable Diffusion, sedangkan proses generasi gambar menggunakan Gemini Nano Banana dilakukan melalui platform *Web Gemini*. Tahap ini mencakup pengujian berbagai variasi *prompt* yang telah dirancang sebelumnya, sehingga setiap model diuji berdasarkan kombinasi *prompt method* yang sama guna memastikan hasil perbandingan yang objektif dan konsisten.

Pada tahap ini, model dieksekusi secara sistematis berdasarkan setiap variasi *prompt method*. Hasil gambar disimpan secara terstruktur dalam direktori dataset dengan metadata yang mencakup jenis model, jenis prompt, parameter, dan waktu eksekusi. Setiap hasil generasi kemudian diekstraksi menggunakan *feature extractor* (CLIP model) untuk memperoleh representasi vektor yang diperlukan pada tahap evaluasi.

3.2.7.4 Evaluation

Tahapan evaluasi dalam penelitian ini dilakukan melalui dua pendekatan utama, yaitu evaluasi objektif dan evaluasi subjektif (*human-based*). Evaluasi objektif dilakukan dengan menghitung metrik kuantitatif untuk menilai kualitas hasil visual yang dihasilkan oleh model *Text-to-Image (T2I)*. Beberapa metrik umum dalam evaluasi model generatif, seperti *Inception Score (IS)* dan *Fréchet Inception Distance (FID)*, pada dasarnya dirancang untuk mengukur kualitas visual berdasarkan kemiripan distribusi gambar hasil generasi terhadap dataset referensi (umumnya ImageNet), di mana IS menekankan pada aspek *diversity* dan *clarity*, sedangkan FID mengukur jarak statistik antara distribusi fitur gambar nyata dan

gambar hasil generasi [43]. Namun, metrik-metrik tersebut tidak secara langsung mengukur apakah gambar yang dihasilkan sesuai dengan makna teks prompt, terutama pada konteks spesifik dan lokal.

Dalam konteks penelitian ini, tujuan utama evaluasi bukanlah menilai kemiripan distribusi gambar dengan *natural images*, melainkan menilai sejauh mana gambar yang dihasilkan mampu merepresentasikan isi prompt berbahasa Indonesia secara semantik dan relevan untuk kebutuhan edukasi anak. Oleh karena itu, *CLIPScore* dipilih sebagai metrik evaluasi objektif karena secara langsung mengukur kesesuaian semantik antara teks dan gambar melalui *cosine similarity* pada *shared embedding space* antara modalitas teks dan visual, dengan rentang nilai 0–1, di mana nilai yang semakin mendekati 1 menunjukkan tingkat kesesuaian semantik yang semakin tinggi. Seluruh proses perhitungan *CLIPScore* dilakukan secara otomatis menggunakan library Python berbasis model CLIP, sehingga menghasilkan evaluasi yang objektif, konsisten, dan dapat direplikasi.

Sementara itu, evaluasi subjektif dilakukan dengan melibatkan panel penilai yang terdiri dari guru sekolah dasar, pengelola Desa Wisata Tigaraksa, serta anak-anak sebagai pengguna akhir. Setiap partisipan diminta memberikan penilaian terhadap hasil visual berdasarkan empat kriteria utama, yaitu: (1) relevansi kontekstual terhadap tema edukatif, (2) akurasi representasi objek, (3) kualitas estetika dan komposisi visual, serta (4) kesesuaian gambar dengan tingkat pemahaman anak-anak sebagai target pengguna. Penilaian dilakukan menggunakan skala Likert 1–5, dan hasilnya dianalisis untuk menggambarkan tingkat penerimaan serta preferensi pengguna terhadap keluaran model.

Dengan mengombinasikan *CLIPScore* sebagai metrik objektif yang berfokus pada *prompt–image alignment* dan human evaluation sebagai penilaian kontekstual dan pedagogis, penelitian ini mampu mengevaluasi performa model generatif secara lebih komprehensif. Pendekatan ini memastikan bahwa evaluasi tidak hanya menilai kualitas visual secara teknis, tetapi juga mempertimbangkan relevansi edukatif, konteks lokal, dan keterpahaman visual, sehingga lebih sesuai dengan

tujuan penerapan teknologi *Text-to-Image* dalam sistem pembelajaran anak di Desa Wisata Tigaraksa.

3.2.7.5 Deployment

Tahap terakhir dari penelitian adalah *deployment*, yaitu penerapan hasil penelitian dalam bentuk website interaktif berbasis AI. Website ini berfungsi sebagai prototipe sistem pembelajaran visual yang memungkinkan pengguna (guru, pengelola, atau siswa) untuk menghasilkan gambar edukatif hanya dengan mengetikkan deskripsi teks.

UI dirancang menggunakan Figma dan Canva agar mudah digunakan oleh pengguna non-teknis, dengan estetika yang cocok untuk anak-anak. Sistem dibangun menggunakan framework Next.js untuk *frontend*, sementara *backend* menggunakan FastAPI yang terhubung dengan API *Text-to-Image* dari Gemini atau Stable Diffusion. Seluruh data hasil generasi disimpan dalam database Supabase.

Tahapan ini juga mencakup user testing, di mana pengguna lokal diminta mencoba sistem secara langsung dan memberikan umpan balik terhadap kecepatan, kemudahan, serta relevansi hasil visual. Tujuannya adalah memastikan bahwa prototipe tidak hanya valid secara teknis, tetapi juga edukatif.

3.3 Teknik Pengumpulan Data

Tahapan pengumpulan data dalam penelitian ini diawali dengan kunjungan ke Desa Wisata Tigaraksa, Kabupaten Tangerang yang menjadi lokasi penelitian. Kegiatan ini dilakukan bersama tim penelitian untuk memahami secara langsung yang diinginkan oleh pengelola desa. Tujuan kunjungan ini adalah untuk mengidentifikasi kebutuhan khususnya dalam penyediaan media pembelajaran visual yang menarik dan relevan bagi anak-anak sekolah dasar.

Dari kegiatan ini diperoleh informasi mengenai apa yang bisa menjadi tema-tema pembelajaran yang sering digunakan dalam kegiatan edukasi anak, seperti pengenalan flora dan fauna lokal serta praktik pertanian. Data hasil observasi dan

wawancara tersebut digunakan untuk menentukan tema-tema utama yang akan dijadikan dasar penyusunan prompt dalam eksperimen *Text-to-Image (T2I)*.

Setelah tema ditentukan, tahap berikutnya adalah perancangan dataset prompt. Setiap tema dikembangkan menjadi beberapa variasi *prompt instance* berdasarkan tiga metode *prompt engineering* yang digunakan dalam penelitian ini, yaitu *Descriptive Prompt, Instruction-Based Prompt dan Compositional Prompt*. Ketiga metode ini dipilih untuk melihat bagaimana perbedaan struktur instruksi dapat memengaruhi hasil gambar yang dihasilkan oleh model.

Selanjutnya, seluruh *prompt* yang telah disusun digunakan dalam proses generasi gambar menggunakan dua model *Text-to-Image*, yaitu *Stable Diffusion* dan *Gemini Nano Banana (Flash)*. Proses generasi dilakukan di lingkungan Jupyter Notebook menggunakan *library Hugging Face Diffusers* untuk Stable Diffusion, serta platform Web Gemini untuk model Gemini. Setiap gambar yang dihasilkan disimpan secara sistematis dalam dataset dengan metadata yang mencakup: tema, jenis prompt, model yang digunakan, serta parameter teknis seperti *guidance scale, inference steps, dan seed value* untuk memastikan replikasi hasil dapat dilakukan secara konsisten.

Dengan demikian, seluruh proses pengumpulan data dalam penelitian ini dilakukan secara berurutan, dimulai dari pengamatan langsung terhadap kebutuhan masyarakat, perumusan tema dan *prompt*, hingga pengumpulan hasil generatif dari model T2I.

3.4 Teknik Analisis Data

Teknik analisis data dalam penelitian ini bertujuan untuk mengevaluasi kualitas hasil visual yang dihasilkan oleh dua model *Text-to-Image (T2I)* yaitu *Stable Diffusion* dan *Gemini Nano Banana (Gemini Flash)* berdasarkan variasi *prompt engineering* yang telah dirancang. Analisis dilakukan dengan menggabungkan pendekatan kuantitatif dan kualitatif agar hasil evaluasi tidak hanya menggambarkan performa teknis model, tetapi juga kesesuaian hasil visual dengan konteks edukatif di Desa Wisata Tigaraksa.

Seluruh proses analisis dilakukan menggunakan *Jupyter Notebook* atau *Google Colab* sebagai lingkungan komputasi utama. Platform ini dipilih karena mendukung eksekusi kode Python secara interaktif dan terintegrasi dengan *library* analisis seperti *Hugging Face Diffusers*, *NumPy*, *Pandas*, *Scikit-learn*, dan *OpenAI CLIP* yang diperlukan untuk menjalankan model generatif serta menghitung metrik evaluasi secara efisien.

Tahapan analisis data diawali dengan proses preprocessing, yaitu penataan dataset hasil generasi gambar berdasarkan metadata yang mencakup tema, metode *prompt* (descriptive, instruction-based, dan compositional) serta model yang digunakan. Struktur ini memudahkan peneliti untuk melakukan analisis perbandingan antara metode dan model secara sistematis. Setelah data siap, dilakukan analisis kuantitatif menggunakan metrik CLIPScore. Selain evaluasi kuantitatif, penelitian ini juga melakukan evaluasi kualitatif (human-based evaluation) untuk menilai aspek visual dari perspektif manusia.

Proses ini melibatkan panel penilai yang terdiri dari guru sekolah dasar, pengelola Desa Wisata Tigaraksa, serta beberapa anak-anak. Setiap partisipan diminta menilai gambar berdasarkan empat aspek utama, yaitu (1) relevansi kontekstual antara gambar dan tema edukatif, (2) akurasi visual terhadap deskripsi teks, (3) estetika dan komposisi visual, serta (4) kesesuaian gambar dengan tingkat pemahaman anak-anak. Penilaian dilakukan menggunakan skala Likert 1–5 untuk memperoleh nilai rata-rata dan persepsi umum dari pengguna.

Dengan demikian, teknik analisis data dalam penelitian ini dirancang untuk memberikan pemahaman menyeluruh terhadap performa model *Text-to-Image* dari dua sisi yaitu teknis model dan persepsi manusia. Pendekatan ini memastikan bahwa hasil penelitian tidak hanya memberikan kontribusi ilmiah dalam bidang *Generative Artificial Intelligence (GenAI)* tetapi juga menghasilkan solusi praktis berupa media pembelajaran berbasis AI yang dapat digunakan secara langsung oleh guru, siswa dan masyarakat Desa Wisata Tigaraksa.



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA