# BAB III
# METODE PENELITIAN

## A. Jenis Penelitian

Penelitian ini merupakan penelitian kuantitatif dengan pendekatan eksperimen komparatif. Pendekatan kuantitatif digunakan karena penelitian ini berlandaskan pada penggunaan data dalam bentuk angka yang dianalisis secara statistik untuk menguji perbedaan antar variabel secara objektif dan sistematis. Metode ini memiliki karakteristik terukur, rasional, dan terstruktur sehingga mampu menghasilkan kesimpulan yang dapat digeneralisasi (Balaka, 2022). Pendekatan eksperimen dipilih karena penelitian ini melibatkan pemberian perlakuan yang berbeda berupa penggunaan dua model Artificial Intelligence, yaitu ChatGPT dan Gemini AI, terhadap objek yang sama, yaitu pertanyaan akademik pada mata kuliah Data Mining, sehingga perbedaan hasil yang diperoleh dapat mencerminkan perbedaan performa masing-masing model.

Dalam penelitian ini, variabel bebas adalah jenis model AI yang digunakan, yaitu ChatGPT dan Gemini AI, sedangkan variabel terikat adalah tingkat kemiripan semantik jawaban model AI terhadap jawaban rujukan yang diukur menggunakan metode *Cosine Similarity*. Variabel dalam penelitian kuantitatif merupakan unit analisis yang diukur secara objektif dan dinyatakan dalam bentuk numerik untuk dianalisis secara statistik (Balaka, 2022). Oleh karena itu, pendekatan kuantitatif digunakan karena hasil penelitian dinyatakan dalam bentuk skor numerik yang memungkinkan dilakukan analisis dan perbandingan performa kedua model AI secara objektif dan sistematis.

## B. Setting Penelitian

Penelitian ini direncanakan dilaksanakan pada awal tahun 2026, yang termasuk dalam semester berjalan tahun akademik 2025/2026. Seluruh proses pengumpulan dan analisis data dilakukan secara daring menggunakan perangkat komputer dan akses internet. ChatGPT dan Gemini AI digunakan melalui platform resmi masing-masing, sedangkan pengolahan data teks dan

perhitungan *Cosine Similarity* dilakukan menggunakan perangkat lunak pendukung pengolahan data teks.

## C. Bahan atau Materi Penelitian

Bahan penelitian dalam penelitian ini berupa 15 pertanyaan akademik yang disusun berdasarkan materi pada mata kuliah *Data Mining*, meliputi konsep dasar *Data Mining*, tahapan *Knowledge Discovery in Databases* (KDD), teknik klasifikasi, clustering, asosiasi, serta topik terkait lainnya. Jumlah 15 pertanyaan dipilih untuk memastikan keterwakilan variasi konsep sekaligus menjaga kestabilan analisis. Dalam penelitian evaluasi berbasis AI, jumlah pertanyaan tidak memiliki batas baku dan dapat disesuaikan dengan tujuan penelitian selama mampu merepresentasikan domain yang diuji. Hal ini sejalan dengan penelitian dalam bidang evaluasi LLM yang menunjukkan bahwa hasil penilaian sangat dipengaruhi oleh desain evaluasi, jenis pertanyaan, dan metode yang digunakan (Kasneci dkk, 2023). Beberapa penelitian sebelumnya bahkan menggunakan jumlah pertanyaan yang lebih sedikit, seperti 13 pertanyaan dalam evaluasi model AI pada domain medis (Strzalkowski dkk, 2024) serta sekitar 10–20 pertanyaan dalam pengujian kemampuan model dalam menjawab pertanyaan akademik (Firdaus dkk, 2024). Oleh karena itu, penggunaan 15 pertanyaan dalam penelitian ini dinilai telah memadai untuk merepresentasikan materi *Data Mining* sekaligus memungkinkan analisis performa model AI dilakukan secara efektif dan terukur.

Untuk memperoleh variasi tingkat kognitif, pertanyaan dalam penelitian ini dikelompokkan ke dalam tiga kategori, yaitu: (1) pertanyaan definisi yang mengukur kemampuan model dalam menjelaskan konsep dasar, (2) pertanyaan konsep yang menilai pemahaman terhadap hubungan antar konsep dalam *Data Mining*, dan (3) pertanyaan analisis yang mengukur kemampuan dalam menguraikan proses atau menyelesaikan permasalahan berbasis konsep. Pengelompokan pertanyaan dilakukan dengan mengadaptasi tingkat kemampuan kognitif dalam *Revised Bloom's Taxonomy* yang membagi proses kognitif ke dalam beberapa tingkat kemampuan, sehingga dapat

digunakan sebagai kerangka untuk mengevaluasi tingkat pemahaman berdasarkan variasi tingkat kognitif, yang dalam penelitian ini diterapkan untuk menilai kemampuan model Artificial Intelligence dalam menjawab pertanyaan dengan tingkat kompleksitas yang berbeda (Anderson & Krathwohl, 2001; Laddha dkk, 2021).

Untuk menjaga *content validity*, jawaban rujukan disusun berdasarkan referensi akademik yang relevan di bidang Data Mining. Validitas isi merupakan aspek penting dalam penelitian karena menunjukkan sejauh mana instrumen mampu merepresentasikan konsep yang diukur secara tepat dan menyeluruh (Masuwai, dkk, 2024) . Oleh karena itu, penyusunan jawaban rujukan dilakukan secara sistematis agar sesuai dengan domain keilmuan yang relevan. Selanjutnya, instrumen penelitian divalidasi oleh dua dosen ahli yang mengampu mata kuliah Data Mining. Proses validasi ini sejalan dengan praktik umum dalam penelitian, di mana penilaian oleh para ahli digunakan untuk memastikan keakuratan, kejelasan, dan kesesuaian isi instrumen sebelum digunakan (Estremera dkk, 2024) . Validator diminta untuk menilai jawaban rujukan berdasarkan beberapa aspek, yaitu kesesuaian materi, kejelasan redaksi, dan ketepatan jawaban, karena aspek-aspek tersebut mencerminkan kualitas instrumen dalam mengukur konstruk yang diteliti.



## D. Alat dan Instrumen Penelitian

Alat dan instrumen yang digunakan dalam penelitian ini meliputi:

1. ChatGPT dan Gemini AI sebagai sumber penghasil jawaban akademik.
2. Perangkat laptop untuk mengakses model AI dan melakukan pengolahan data.
3. *Google Colaboratory* (*Google Colab*) sebagai lingkungan komputasi untuk pengolahan data teks.
4. Microsoft Excel sebagai alat bantu pengelolaan dan rekapitulasi data.
5. Literatur Data Mining yang digunakan sebagai sumber rujukan dalam penyusunan pertanyaan dan jawaban rujukan (*ground truth*).
6. *Streamlit* sebagai *framework* berbasis Python untuk membangun sistem visualisasi hasil penelitian secara interaktif berbasis web.

Instrumen utama dalam penelitian ini adalah daftar pertanyaan akademik Mata Kuliah *Data Mining* yang sama untuk kedua model AI.

**E. Teknik Pengumpulan Data**

Teknik pengumpulan data dalam penelitian ini dilakukan melalui beberapa tahapan sebagai berikut:

1. Penyusunan Pertanyaan Akademik: Peneliti menyusun 15 pertanyaan berdasarkan materi Data Mining dengan variasi tingkat kognitif berupa definisi, konsep, dan analisis.
2. Penyusunan Jawaban Rujukan: Peneliti menyusun jawaban rujukan berdasarkan referensi akademik yang relevan di bidang Data Mining untuk digunakan sebagai acuan penilaian.
3. Validasi Jawaban Rujukan: Jawaban rujukan divalidasi oleh dosen ahli untuk memastikan kesesuaian dan ketepatan isi.
4. Pengumpulan Data Jawaban AI: Pertanyaan diajukan kepada ChatGPT dan Gemini AI untuk memperoleh jawaban dari masing-masing model.
5. Dokumentasi Data: Seluruh data pertanyaan dan jawaban didokumentasikan dalam bentuk teks untuk analisis selanjutnya.

**F. Teknik Analisis Data**



Analisis data dalam penelitian ini dilakukan melalui beberapa tahapan yang sistematis untuk mengukur tingkat kemiripan jawaban yang dihasilkan oleh model AI terhadap jawaban rujukan akademik. Tahap pertama adalah pra-pemrosesan teks (*preprocessing*) yang bertujuan untuk menyiapkan dan meningkatkan kualitas data sebelum dilakukan analisis lebih lanjut. Proses ini meliputi normalisasi teks, penghapusan karakter yang tidak relevan, serta penyamaan format penulisan. Tahap *preprocessing* penting dilakukan karena kualitas teks mempengaruhi hasil representasi *embedding*, di mana data yang lebih bersih dan konsisten akan menghasilkan representasi semantik yang lebih akurat dalam pengukuran kemiripan (Koto dkk, 2020).

Tahap selanjutnya adalah representasi teks menggunakan model berbasis *transformer*, yaitu *Sentence-BERT* (SBERT). Pada tahap ini, jawaban yang dihasilkan oleh model AI dan jawaban rujukan diubah menjadi vektor

*embedding* yang merepresentasikan makna semantik pada tingkat kalimat. Pendekatan berbasis *embedding* kontekstual memungkinkan sistem mengenali kesamaan makna antar teks meskipun memiliki perbedaan struktur kalimat. Penggunaan SBERT dalam tugas *semantic textual similarity* telah terbukti efektif dalam merepresentasikan hubungan semantik antar kalimat (Kasneci dkk, 2023; Reimers & Gurevych, 2019).

Setelah diperoleh representasi vektor, dilakukan pengukuran tingkat kemiripan menggunakan metode *Cosine Similarity*. Metode ini digunakan untuk menghitung kesamaan antara dua vektor berdasarkan nilai cosinus sudut di antara keduanya, sehingga mampu mengukur kedekatan makna antar teks secara numerik. Secara matematis, *Cosine Similarity* dirumuskan sebagai berikut:

$$\cos(\theta) = \frac{A \cdot B}{\|A\| . \|B\|} \tag{3.1}$$

di mana $A$ dan $B$ merupakan vektor yang dibandingkan, $A \cdot B$ adalah *dot product*, serta $\| A \|$dan $\| B \|$adalah norma dari masing-masing vektor. Nilai kemiripan berada pada rentang 0 hingga 1, di mana nilai yang mendekati 1 menunjukkan tingkat kemiripan semantik yang tinggi.

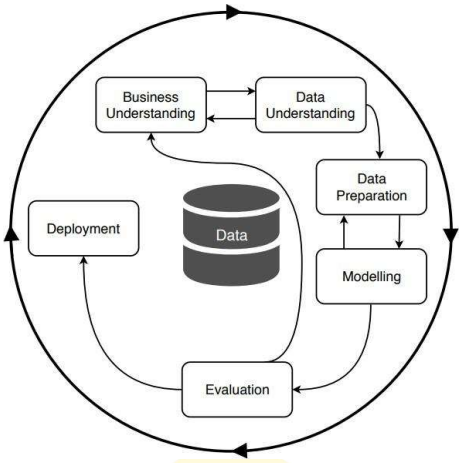
Dalam penelitian ini, nilai kemiripan semantik dari tiga kali percobaan pada setiap pertanyaan dirata-ratakan untuk memperoleh nilai yang lebih representatif serta mengurangi variasi akibat ketidakstabilan keluaran model. Selanjutnya, untuk menguji perbedaan performa antara ChatGPT dan Gemini AI, digunakan uji Mann-Whitney karena jumlah data relatif kecil dan tidak diasumsikan berdistribusi normal. Uji ini merupakan metode non-parametrik untuk membandingkan dua kelompok independen tanpa asumsi normalitas (Dehaene dkk, 2021). Kriteria pengujian didasarkan pada nilai signifikansi (*p-value*), di mana jika nilai $p < 0{,}05$ maka terdapat perbedaan yang signifikan antara kedua model.

Untuk mendukung analisis tersebut, dirumuskan hipotesis penelitian sebagai berikut:

- H0 : Tidak terdapat perbedaan signifikan nilai kemiripan semantik antara ChatGPT dan Gemini AI.

- H1 :Terdapat perbedaan signifikan nilai kemiripan semantik antara ChatGPT dan Gemini AI.

## G. Alur Penelitian (*Flowchart*)




Gambar 3 1 Proses Metode CRISP-DM

Sumber: (Martínez-Plumed dkk., 2019)

Gambar 3.1 menunjukkan alur penelitian yang digunakan untuk membandingkan kualitas jawaban akademik yang dihasilkan oleh ChatGPT dan Gemini AI pada mata kuliah Data Mining menggunakan metode Cosine Similarity. Alur penelitian ini mengacu pada pendekatan CRISP-DM yang terdiri dari enam tahapan utama, yaitu *business understanding, data understanding, data preparation, modeling, evaluation,* dan *deployment*. Adapun tahapan penelitian yang dilakukan adalah sebagai berikut:

1. *Business Understanding*

   Tahap *business understanding* merupakan tahap awal yang bertujuan untuk memahami permasalahan dan menetapkan tujuan penelitian. Permasalahan dalam penelitian ini adalah bagaimana mengukur tingkat kedekatan semantik antara jawaban yang dihasilkan oleh model AI dengan jawaban rujukan akademik, serta membandingkan performa ChatGPT dan Gemini AI dalam menjawab pertanyaan pada mata kuliah Data Mining. Pada tahap ini juga ditentukan pendekatan yang digunakan, yaitu metode Cosine Similarity berbasis representasi embedding menggunakan Sentence-BERT (SBERT) sebagai solusi untuk mengukur kemiripan

semantik antar teks. Dengan demikian, tahap ini menjadi dasar dalam menentukan arah dan tujuan keseluruhan penelitian.

2. *Data Understanding*

   Tahap *data understanding* bertujuan untuk memahami karakteristik data yang digunakan dalam penelitian. Data yang digunakan terdiri dari pertanyaan akademik sebanyak 15 soal yang mencakup materi Data Mining seperti konsep dasar, proses Knowledge Discovery in Databases (KDD), teknik klasifikasi, *clustering*, asosiasi, serta topik terkait lainnya. Selain itu, data juga mencakup jawaban rujukan yang disusun berdasarkan referensi akademik yang relevan di bidang Data Mining sebagai ground truth, serta jawaban yang dihasilkan oleh ChatGPT dan Gemini AI. Pada tahap ini dilakukan pula pengelompokan pertanyaan ke dalam kategori definisi, konsep, dan analisis untuk merepresentasikan variasi tingkat kognitif. Selanjutnya, dilakukan proses validasi terhadap pertanyaan dan jawaban rujukan oleh dua dosen ahli dengan menilai kesesuaian materi, kejelasan redaksi, dan ketepatan jawaban.

3. *Data Preparation*

   Tahap *data preparation* bertujuan untuk menyiapkan data agar siap digunakan dalam proses pemodelan. Pada tahap ini dilakukan pengumpulan jawaban dari ChatGPT dan Gemini AI sebanyak tiga kali percobaan dengan menggunakan format *prompt* yang sama, di mana *prompt* dirancang untuk menghasilkan jawaban dalam bentuk paragraf guna menjaga konsistensi format. Jawaban yang dihasilkan kemudian digunakan secara langsung tanpa dilakukan penyuntingan, sehingga diperoleh keluaran asli (*raw output*) dari masing-masing model. Setelah itu, dilakukan proses *preprocessing* teks yang meliputi normalisasi teks, penghapusan karakter yang tidak relevan, penghapusan spasi berlebih, serta penyamaan format penulisan, dengan tujuan meningkatkan kualitas dan konsistensi data tanpa mengubah makna kalimat, sehingga dapat menghasilkan representasi semantik yang lebih akurat.

4. *Modeling*

Tahap *modeling* dilakukan untuk merepresentasikan teks ke dalam bentuk numerik serta menghitung tingkat kemiripan semantik antar teks. Pada tahap ini, teks yang telah melalui *preprocessing* direpresentasikan ke dalam bentuk vektor embedding menggunakan model Sentence-BERT (SBERT). Representasi ini memungkinkan sistem untuk menangkap makna semantik pada tingkat kalimat, meskipun terdapat perbedaan struktur atau susunan kata. Selanjutnya, dilakukan perhitungan nilai kemiripan menggunakan metode *Cosine Similarity* antara jawaban model AI dan jawaban rujukan. Nilai yang dihasilkan berada pada rentang 0 hingga 1, di mana nilai yang mendekati 1 menunjukkan tingkat kemiripan semantik yang semakin tinggi.

5. *Evaluation*

Tahap *evaluation* bertujuan untuk menganalisis dan mengevaluasi hasil yang diperoleh dari proses pemodelan. Pada tahap ini, nilai *Cosine Similarity* digunakan sebagai ukuran kedekatan semantik antara jawaban model AI dan ground truth. Nilai similarity berada pada rentang 0 hingga 1, di mana nilai yang semakin mendekati 1 menunjukkan bahwa makna kedua teks semakin serupa, sedangkan nilai yang semakin mendekati 0 menunjukkan bahwa kedua teks memiliki tingkat kesamaan makna yang semakin rendah. Nilai dari tiga kali percobaan pada setiap pertanyaan kemudian dirata-ratakan untuk memperoleh hasil yang lebih representatif. Selanjutnya, dilakukan analisis untuk membandingkan performa ChatGPT dan Gemini AI berdasarkan nilai kemiripan semantik yang dihasilkan. Untuk memperkuat hasil analisis, dilakukan uji statistik Mann-Whitney guna mengetahui apakah terdapat perbedaan yang signifikan antara kedua model. Pengujian ini didasarkan pada nilai signifikansi (*p-value*), di mana jika nilai $p < 0,05$ maka terdapat perbedaan yang signifikan.

6. *Deployment*

Tahap *deployment* merupakan tahap implementasi hasil penelitian ke dalam bentuk sistem sederhana berbasis web menggunakan Streamlit. Pada tahap ini, hasil perhitungan kemiripan semantik antara jawaban model AI dan jawaban rujukan akademik diintegrasikan ke dalam sistem. Sistem tersebut digunakan untuk menampilkan hasil penelitian secara terstruktur sehingga memudahkan proses analisis dan interpretasi data. Selain itu, hasil yang ditampilkan juga dapat membantu dalam melihat perbandingan performa antara ChatGPT dan Gemini AI secara lebih jelas. Dengan demikian, tahap deployment berfungsi sebagai media visualisasi hasil agar lebih mudah dipahami.