

BAB I

PENDAHULUAN

1.1 Latar Belakang

Evaluasi pembelajaran menjadi salah satu bagian penting dalam proses pendidikan tinggi karena berfungsi untuk melihat sejauh mana mahasiswa telah mencapai kompetensi yang diharapkan, terutama dalam aspek berpikir tingkat tinggi. Salah satu bentuk penilaian atau asesmen yang sering digunakan untuk mengukur kemampuan analitis, kritis, dan sintesis mahasiswa adalah soal uraian. Menurut temuan Rokhman et al., (2025), Penilaian berbasis uraian memiliki keunggulan dalam menggali kemampuan mahasiswa dalam mengaitkan konsep, menyusun argumen, serta menghasilkan jawaban yang lebih kreatif dibandingkan soal pilihan ganda. Karena itu, asesmen jenis ini memiliki peran penting dalam memastikan mahasiswa tidak hanya menghafal, tetapi benar-benar memahami dan mampu mengaplikasikan pengetahuannya dalam situasi nyata.

Walau demikian, praktik penilaian uraian secara manual masih menyimpan sejumlah kendala. Berdasarkan wawancara mendalam yang dilakukan pada tahun 2025 dengan Ibu Luluk Mauli Diana, S.Si., M.Pd., selaku dosen pengampu mata kuliah Strategi Pembelajaran Kejuruan di Program Studi Pendidikan Informatika Universitas Trunodjoyo Madura, ditemukan permasalahan nyata dalam proses evaluasi,. Beliau mengungkapkan bahwa penilaian jawaban uraian memiliki kendala utama berupa tingkat subjektivitas yang tinggi, proses koreksi yang memakan waktu lama (berkisar 3-4 jam per kelas), serta keterlambatan dalam memberikan umpan balik kepada mahasiswa karena keterbatasan waktu luang

dosen. Hasil ini sejalan dengan Azzaroh et al. (2025) yang menyebutkan bahwa penilaian manual kerap tidak konsisten antar penilai dan memperlambat proses refleksi mahasiswa terhadap hasil belajarnya. Kesenjangan antara kebutuhan akan penilaian yang cepat dan adil dengan realitas koreksi manual yang lambat serta bias inilah yang menjadi pendorong utama dilakukannya penelitian ini.

Sebagai solusi atas permasalahan tersebut, diperlukan upaya pengembangan sistem yang mampu melakukan koreksi jawaban uraian secara otomatis. Pengembangan sistem dalam konteks ini dipahami sebagai sebuah proses rancang bangun teknologi yang bertujuan untuk menerjemahkan kebutuhan evaluasi ke dalam bentuk aplikasi yang fungsional (Dermawan & Hartini, 2017). Teknologi yang digunakan untuk menjawab tantangan ini adalah *Automated Essay Scoring* (AES). Sistem AES bekerja dengan memanfaatkan kecerdasan buatan untuk mengevaluasi kualitas linguistik, sintaksis, dan semantik dari sebuah tulisan agar komputer dapat memberikan skor yang memiliki tingkat kepercayaan tinggi dan mendekati penilaian manusia (Yakubu et al., 2024).

Sistem yang dikembangkan dalam penelitian ini dirancang dengan platform berbasis web. Pemilihan aplikasi berbasis web didasarkan pada pertimbangan kemudahan akses, di mana pengguna dapat menjalankan aplikasi dari berbagai perangkat tanpa harus melakukan instalasi perangkat lunak tambahan (Rifai & Yuniar, 2019). Untuk memastikan sistem berjalan stabil dan aman, pengembangan ini menggunakan kerangka kerja (*framework*) Laravel yang memiliki struktur kode terorganisir serta manajemen basis data yang handal (Nashir Nasution & Maulana, 2024). Keunggulan dari sistem berbasis web adalah kemampuannya dalam memfasilitasi interaksi antara dosen dan mahasiswa secara lebih dinamis, efektif,

dan efisien melalui jaringan internet (Sastra, 2017).

Kebaruan utama dalam penelitian ini terletak pada sifat sistem yang terintegrasi dengan model bahasa besar generasi terbaru. Istilah terintegrasi merujuk pada hubungan fungsional antara sistem web dengan mesin pengolah bahasa alami melalui antarmuka pemrograman aplikasi atau API (Thaib, 2024). Model kecerdasan buatan yang dipilih adalah Gemini 2.0 *Flash*. Pemilihan Gemini 2.0 *Flash* didasarkan pada arsitektur teknisnya yang sangat unggul dibandingkan model AI lainnya maupun metode pencocokan teks konvensional. Gemini 2.0 *Flash* dibangun dengan arsitektur *Sparse Mixture-of-Experts* (MoE) yang bekerja dengan cara mengaktifkan hanya sebagian kecil parameter yang paling relevan untuk setiap *input*, sehingga proses pengolahan data menjadi jauh lebih cepat dan hemat biaya komputasi (Rahman et al., 2025).

1.2 Rumusan Masalah

Berdasarkan gambaran masalah yang telah diuraikan, rumusan masalah dalam penelitian ini dapat dirinci sebagai berikut:

1. Bagaimana mengembangkan sistem ujian berbasis web yang terintegrasi dengan Model Gemini?
2. Bagaimana pengujian kelayakan sistem ujian yang terintegrasi dengan Model Gemini terhadap penilaian jawaban uraian mahasiswa pada mata kuliah Strategi Pembelajaran Kejuruan di prodi Pendidikan Informatika dengan pengujian ISO 25010?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah diajukan, tujuan dari penelitian ini adalah sebagai berikut:

1. Mengembangkan sistem berbasis web yang terintegrasi dengan Model Gemini
2. Menguji kelayakan sistem ujian yang terintegrasi dengan Model Gemini dalam menilai jawaban uraian mahasiswa.

1.4 Spesifikasi Produk yang Diharapkan

Produk yang dikembangkan dalam penelitian ini berupa sistem ujian berbasis *website* yang terintegrasi dengan model bahasa besar Gemini 2.0 *Flash*. Sistem ini dirancang untuk membantu dosen dalam melakukan penilaian jawaban uraian mahasiswa secara cepat, objektif, dan konsisten, serta memberikan umpan balik yang lebih terukur bagi mahasiswa. Spesifikasi produk yang diharapkan adalah sebagai berikut:

1.4.1 Platform Berbasis Web

Sistem dapat diakses melalui browser pada perangkat komputer maupun gawai, sehingga memudahkan dosen dan mahasiswa dalam menggunakannya tanpa perlu instalasi tambahan.

1.4.2 Manajemen Ujian

- a. Dosen dapat membuat soal uraian dengan berbagai tingkat kesulitan.
- b. Mahasiswa dapat mengikuti ujian secara daring melalui akun masing-masing.
- c. Sistem mendukung pengaturan waktu ujian, jumlah soal, serta pengelolaan peserta.

1.4.3 Integrasi Model Gemini 2.0 *Flash*

- a. Sistem terhubung dengan *Large Language Model* Gemini 2.0 *Flash* untuk melakukan analisis semantik dan sintaksis jawaban mahasiswa.
- b. Model mampu memberikan skor otomatis berdasarkan indikator penilaian yang ditetapkan.

1.4.4 Fitur Penilaian Otomatis Jawaban Uraian

- a. Analisis semantik jawaban dilakukan dengan cara mengukur kesesuaian makna antara jawaban mahasiswa dan rubrik, bukan sekadar pencocokan kata kunci.
- b. Hasil penilaian ditampilkan dalam bentuk skor kuantitatif beserta umpan balik deskriptif.

1.4.5 Dashboard Dosen dan Mahasiswa

- a. Dosen: dapat memantau hasil penilaian, mengunduh laporan nilai, serta melakukan validasi terhadap hasil penilaian otomatis.
- b. Mahasiswa: dapat melihat hasil penilaian dan umpan balik secara langsung setelah ujian selesai.

1.4.6 Keamanan dan Privasi Data

- a. Setiap pengguna memiliki akun tersendiri dengan autentikasi yang aman.
- b. Data jawaban mahasiswa disimpan di *database* dengan enkripsi agar terjamin kerahasiaannya.

1.4.7 Teknologi yang Digunakan

- a. *Backend*: Laravel *Framework* untuk logika aplikasi.
- b. CMS: Voyager untuk manajemen konten.
- c. *Frontend*: Bootstrap untuk tampilan antarmuka responsif.
- d. *Database*: MySQL untuk penyimpanan data soal, jawaban, dan nilai.

1.4.8 Kelayakan Produk

Produk akan diuji kelayakannya melalui validasi ahli media, ahli materi, serta uji coba terbatas pada mahasiswa Program Studi Pendidikan Informatika. Uji coba meliputi aspek fungsionalitas, kemudahan penggunaan (*usability*), dan kesesuaian

penilaian dengan standar evaluasi pembelajaran.

Dengan spesifikasi tersebut, produk akhir diharapkan mampu menjadi solusi praktis bagi permasalahan penilaian manual jawaban uraian, serta memberikan kontribusi nyata dalam mendukung transformasi digital di bidang evaluasi pembelajaran.

1.5 Batasan Penelitian Pengembangan

Dalam penelitian pengembangan ini, produk yang dihasilkan berupa sistem ujian berbasis *website* yang terintegrasi dengan model Gemini 2.0 *Flash*. Untuk menjaga fokus penelitian dan memudahkan proses pengembangan, penelitian ini dibatasi pada hal-hal berikut:

1.5.1 Ruang Lingkup Materi Ujian

- a. Penelitian hanya diterapkan pada mata kuliah *Strategi Pembelajaran Kejuruan* di Program Studi Pendidikan Informatika Universitas Trunodjoyo Madura.
- b. Soal ujian yang digunakan berbentuk uraian terbuka sesuai dengan capaian pembelajaran mata kuliah tersebut.

1.5.2 Fokus Pengembangan Sistem

- a. Sistem difokuskan pada fungsi koreksi otomatis jawaban uraian dengan integrasi model Gemini 2.0 *Flash*.
- b. Penilaian diarahkan pada aspek kesesuaian isi, kelengkapan jawaban, dan struktur argumen, bukan pada aspek non-verbal seperti diagram, kode program kompleks, atau simbol matematis tingkat lanjut.

1.5.3 Subjek Uji Coba

- a. Uji coba sistem hanya dilakukan pada lingkup terbatas, yaitu dosen

pengampu mata kuliah, validator ahli dan mahasiswa Pendidikan Informatika sebagai pengguna.

b. Jumlah responden dibatasi sesuai dengan kebutuhan validasi dan uji coba terbatas (bukan uji coba berskala luas).

1.5.4 Teknologi yang Digunakan

a. Sistem dibangun dengan *framework* Laravel sebagai backend, Voyager sebagai CMS, Bootstrap untuk tampilan antarmuka, dan MySQL sebagai basis data.

b. Integrasi LLM menggunakan Gemini 2.0 *Flash* dilakukan melalui API, dengan batasan infrastruktur sesuai kapasitas server penelitian.

Dengan adanya batasan-batasan tersebut, penelitian ini lebih terarah pada pengembangan prototipe sistem ujian yang layak diuji coba, sehingga dapat memberikan kontribusi sesuai dengan tujuan penelitian yang telah ditetapkan.

1.6 Pentingnya Penelitian Pengembangan

Penelitian pengembangan ini memiliki tingkat kepentingan yang tinggi, baik secara teoretis maupun praktis, karena berupaya menjawab kesenjangan antara kebutuhan akan evaluasi pembelajaran yang cepat, objektif, dan konsisten dengan keterbatasan metode manual yang masih dominan digunakan di perguruan tinggi. Pentingnya penelitian ini dapat dijelaskan sebagai berikut:

1.6.1 Penting dari Segi Teoretis

a. Penelitian ini memperkaya literatur mengenai penerapan kecerdasan buatan, khususnya *Large Language Model* (LLM), dalam bidang evaluasi pendidikan.

b. Penelitian ini mendukung penguatan kajian tentang *Automated Essay Scoring* (AES) berbasis NLP, yang masih relatif baru dalam konteks pendidikan tinggi

di Indonesia.

1.6.2 Penting dari Segi Praktis

a. Bagi dosen, sistem ini dapat mengurangi beban kerja dalam mengoreksi jawaban uraian, meningkatkan konsistensi penilaian, dan mempercepat pemberian umpan balik kepada mahasiswa.

b. Bagi mahasiswa, sistem ini memungkinkan mereka memperoleh hasil penilaian yang lebih cepat, transparan, dan objektif, sehingga dapat segera melakukan refleksi serta perbaikan pemahaman materi.

Berdasarkan paparan tersebut, penelitian pengembangan ini tidak hanya mendesak untuk dilakukan karena adanya kebutuhan praktis di lapangan, tetapi juga penting dalam konteks keilmuan dan kebijakan pendidikan di era transformasi digital. Dengan demikian, hasil penelitian ini diharapkan mampu memberikan dampak nyata, baik dalam pengembangan ilmu pengetahuan maupun dalam peningkatan kualitas evaluasi pembelajaran di perguruan tinggi.

1.7 Definisi Istilah

Untuk menghindari adanya perbedaan penafsiran terhadap istilah-istilah yang digunakan dalam penelitian ini, maka perlu diberikan definisi operasional sebagai berikut:

1.7.1 *Artificial Intelligence (AI)*

Artificial Intelligence (AI) atau kecerdasan buatan didefinisikan sebagai salah satu bidang dalam dunia komputasi yang diciptakan agar dapat berfungsi sebagai sesuatu yang mampu berpikir layaknya cara berpikir manusia, sesuai dengan (Thaib, 2024). AI merupakan teknologi yang memungkinkan sistem komputer untuk melakukan tugas yang biasanya memerlukan kecerdasan manusia, seperti

pembelajaran dan pengambilan keputusan (Thaib, 2024).

1.7.2 Sistem Penilaian Jawaban Uraian Otomatis (*Automated Essay Scoring/AES*)

Sistem Penilaian Jawaban Uraian Otomatis, atau *Automated Essay Scoring* (AES) atau *Automated Essay Grading* (AEG), adalah pendekatan berbasis teknologi yang digunakan untuk mengevaluasi esai tertulis (Yakubu et al., 2024). Sistem ini memanfaatkan model statistik, pembelajaran mesin (*machine learning*), dan *Natural Language Processing* (NLP) untuk menilai berbagai aspek uraian, seperti relevansi, koherensi, pengembangan ide (Susilawati et al., 2025; Yakubu et al., 2024). Sistem AES meniru penilaian manusia dengan mengevaluasi kualitas linguistik, termasuk tata bahasa, sintaksis, dan semantik (Yakubu et al., 2024).

1.7.3 *Natural Language Processing* (NLP)

Natural Language Processing (NLP) adalah cabang dari Kecerdasan Buatan (AI) yang berfokus pada interaksi antara komputer dan bahasa manusia (Ratnayanti et al., 2023). Teknik NLP diterapkan dalam penilaian bahasa untuk menganalisis dan memahami teks atau ucapan, memungkinkan pemahaman otomatis dan pemberian umpan balik (Ratnayanti et al., 2023). Dalam konteks penilaian uraian, model NLP berbasis *deep learning*, seperti BERT (*Bidirectional Encoder Representations from Transformers*), digunakan untuk menilai kesamaan makna (semantik) antar teks (Azzaroh et al., 2025).

1.7.4 Sistem Ujian Online Berbasis Web

Sistem ujian *online* adalah sebuah sistem terintegrasi yang dirancang untuk menyediakan dan menyelenggarakan ujian secara lebih cepat dan efektif, yang bertujuan untuk mengetahui mutunya (Thaib, 2024). Sistem ini memanfaatkan

perangkat keras dan perangkat lunak komputer, prosedur manual, dan basis data untuk menyelenggarakan seluruh proses ujian. Sistem ujian online dapat meningkatkan aksesibilitas ujian dan menyediakan fleksibilitas waktu bagi peserta ujian, serta memungkinkan pengawasan otomatis dan penilaian otomatis untuk jenis soal tertentu

1.7.5 Soal Uraian

Soal uraian adalah bentuk tes yang disusun dalam bentuk pertanyaan terstruktur yang memerlukan jawaban berupa penjelasan atau uraian kata-kata (Thaib, 2024; Yulyantari, 2017). Soal uraian adalah jenis evaluasi di mana pilihan jawaban tidak disediakan, dan peserta didik harus menyusun serta mengorganisasikan sendiri jawaban tiap pertanyaan itu dengan bahasa sendiri (Yulyantari, 2017). Akibatnya, jawaban dari peserta didik sangat bervariasi sesuai dengan pemikiran masing-masing.

1.7.6 Model *Waterfall*

Model *Waterfall*, atau dikenal sebagai *Linear Sequential Model* atau siklus hidup klasik, adalah pendekatan yang sistematis dan berurutan dalam pengembangan perangkat lunak (Thaib, 2024). Prosesnya dimulai dari pengumpulan kebutuhan pengguna, kemudian dilanjutkan dengan tahapan perencanaan, permodelan, konstruksi, hingga penyebaran sistem kepada pengguna (Thaib, 2024). Metode ini menekankan bahwa setiap tahap harus diselesaikan secara lengkap sebelum melanjutkan ke tahap berikutnya, dan tidak ada kemungkinan untuk kembali ke tahap sebelumnya setelah tahap tersebut terlewati.

1.7.7 Gemini 2.0 *Flash*

Gemini 2.0 Flash merupakan varian model bahasa besar (Large Language Model)

multimodal generasi terbaru yang dirancang untuk mengoptimalkan kecepatan inferensi dan latensi rendah (Gemini Team, 2025). Secara teknis, model ini mengadopsi arsitektur *Sparse Mixture-of-Experts* (MoE) yang memungkinkan pemrosesan informasi secara efisien melalui pengaktifan parameter relevan, serta mendukung jendela konteks (*context window*) berskala besar untuk memahami teks akademik yang kompleks (Gemini Team, 2025).

1.7.8 Terintegrasi

Terintegrasi didefinisikan sebagai kondisi di mana berbagai sub-sistem atau komponen perangkat lunak yang berbeda saling berhubungan dan berkolaborasi melalui API dalam satu kesatuan fungsional untuk menyelesaikan masalah tertentu (Rifai & Yuniar, 2019; Thaib, 2024).