

BAB I

PENDAHULUAN

1.1 Latar Belakang

Evaluasi akademik merupakan fondasi kritis dalam sistem pendidikan modern yang menentukan kualitas pembelajaran dan pengembangan intelektual (Yakub Iskandar dkk., 2024). Penilaian akademik tidak hanya penting untuk mengukur prestasi akademik tetapi juga untuk membangun strategi untuk pengembangan kompetensi mahasiswa secara menyeluruh (Ahmad Qurtubi dkk., 2023). Namun, penerapannya pada penilaian berbasis teks, terutama esai, menjadi rumit karena adanya kebutuhan untuk menjaga penilaian yang objektif dan benar-benar akurat. (Chamidah dkk., 2021).

Esai dipilih sebagai alat penilaian karena bisa menunjukkan kemampuan mahasiswa dalam menganalisis, menyusun argumen yang runtut, dan memahami suatu konsep secara mendalam. Tidak seperti soal pilihan ganda, esai membuat mahasiswa harus menyusun pemikiran dengan teratur, menghubungkan berbagai konsep dari beberapa bidang, serta membuktikan pendapatnya dengan alasan yang logis. Proses ini menggambarkan bagaimana mereka menerapkan pengetahuan dalam situasi nyata (Ayu Pradani dkk., 2023).

Penilaian esai manual menghadapi dua masalah utama yaitu inkonsistensi akibat subjektivitas penilai (Mochamad dkk., 2024), dan beban waktu tinggi (Putra Pane dkk., 2024). Salah satu solusi umum untuk mengatasi masalah ini adalah meningkatkan tenaga penilai. Namun, hal ini berdampak pada peningkatan biaya operasional dan ketidakseragaman standar penilaian (Putri dkk., 2022). Muslich

dikutip dalam (Ayu Pradani dkk., 2023) menyatakan bahwa dengan kunci jawaban, variasi ekspresi linguistik dan struktur esai yang unik menyebabkan penilaian manual rentan terhadap inkonsistensi.

Cosine Similarity diadopsi dalam penelitian ini karena kemampuannya mengkuantifikasi kesamaan semantik antar teks secara matematis, meminimalkan bias subjektif manusia (Kinanti & Qoiriah, 2020). Metode ini telah terbukti efektif dalam berbagai aplikasi penilaian otomatis, termasuk penilaian esai, karena kemampuannya untuk menangkap kemiripan tekstual. Namun, meskipun *Cosine Similarity* memberikan dasar yang kuat untuk penilaian otomatis, ketergantungan pada metode pemrosesan teks sederhana dapat membatasi kemampuan sistem dalam menangkap nuansa bahasa dan makna semantik yang lebih dalam (Dhini dkk., 2023).

Penelitian terdahulu menunjukkan perkembangan metode penilaian esai otomatis berbasis *text similarity*, namun mengidentifikasi celah kritis. (Hidayatul Ula & Yuliyanti, 2024) membuktikan efektivitas *Cosine Similarity* dalam sistem web, tetapi hasilnya bergantung pada kualitas *dataset* dan gagal menangkap kompleksitas jawaban esai. Hasil penelitian terdahulu (Rosnelly dkk., 2021) mencapai nilai akurasi 89,5% dengan *Cosine Similarity*, namun nilai akurasi turun drastis pada jawaban dengan struktur kalimat tidak baku atau penggunaan sinonim di luar basis data. Penelitian lain (Lahitani, 2022) mengembangkan penilaian esai otomatis berbasis similaritas teks, tetapi mensyaratkan jumlah kata jawaban harus sama persis dengan kunci batasan yang tidak realistis dalam konteks esai argumentatif. Penelitian berikutnya (Ahmad et al., 2020) berhasil meningkatkan

nilai akurasi hingga 90,66% dengan kombinasi *stemming* Nazief & Adriani dan *frequency accuracy*, tetapi metode ini terbatas pada pencocokan kata kunci tanpa mempertimbangkan koherensi semantik. Sementara itu (Kinanti & Qoiriah, 2020) mengusulkan SynSemSim untuk esai Bahasa Indonesia, tetapi nilai akurasi hanya 70,44% akibat tantangan struktur kalimat dan variasi kosakata. *Gap* utama dari seluruh studi tersebut adalah: (1) ketergantungan berlebihan pada struktur teks statis, (2) Ketidakmampuan untuk mengenali variasi penggunaan bahasa dan menafsirkan makna yang berubah sesuai konteks, serta (3) absennya mekanisme umpan balik kualitatif untuk pengembangan kompetensi mahasiswa.

Upaya untuk mengatasi tantangan ini seperti (1) ketergantungan berlebihan pada struktur teks statis, (2) Ketidakmampuan untuk mengenali variasi penggunaan bahasa dan menafsirkan makna yang berubah sesuai konteks, serta (3) absennya mekanisme umpan balik kualitatif untuk pengembangan kompetensi mahasiswa, penelitian ini menggabungkan *Cosine Similarity* dan *Model* (LLM) bertujuan untuk menghadirkan terobosan dalam sistem penilaian esai otomatis yang mampu mengatasi keterbatasan metode konvensional (Ishida dkk., 2023). Dalam pendekatan ini, *Cosine Similarity* akan berperan sebagai metode awal untuk menghitung kesamaan semantik antara jawaban esai mahasiswa dan kunci jawaban, mengukur kesesuaian kata kunci dan struktur dasar jawaban. Selanjutnya *Model* (LLM) berperan untuk analisis mendalam terhadap konten esai, mengeksplorasi nuansa makna, konteks, dan kedalaman argumen yang tidak dapat ditangkap hanya melalui perhitungan kesamaan kata. LLM tidak hanya akan memberikan skor numerik, tetapi juga menghasilkan umpan balik menyeluruh, seperti Dengan

memberikan catatan yang rinci tentang alasan sebuah esai mendapatkan nilai tertentu, LLM dapat menjelaskan bagian mana yang kuat dari argumen, bagaimana alur logikanya, seberapa dalam analisis yang diberikan, serta bagian-bagian mana yang masih perlu dikembangkan secara lebih spesifik (Kostic dkk., 2024). Melalui kombinasi kedua metode ini, diharapkan hasil penilaian akan lebih konsisten, objektif, dan informatif, sehingga dapat meningkatkan kepercayaan pengajar terhadap sistem evaluasi otomatis di lingkungan pendidikan modern dan memberikan panduan konstruktif bagi mahasiswa untuk meningkatkan keterampilan akademik mereka.

1.2 Rumusan Masalah

1. Bagaimana mengembangkan model penilaian esai otomatis berbasis kombinasi *Cosine Similarity* dan *Model* (LLM)?
2. Bagaimana mengukur performa model *Cosine Similarity* dan *Model* (LLM) dalam menilai jawaban esai mahasiswa?
3. Bagaimana mengevaluasi efektivitas model hibrid *Cosine Similarity* dan *Model* (LLM) dalam meningkatkan akurasi dan konsistensi penilaian esai dibandingkan penilaian manual?

1.3 Tujuan Penelitian

1. Mengembangkan model penilaian esai otomatis berbasis kombinasi *Cosine Similarity* dan *Model* (LLM).
2. Mengukur performa model *Cosine Similarity* dan *Model* (LLM) dalam proses penilaian esai mahasiswa untuk mengetahui tingkat akurasi dan keandalan masing-masing model.

3. Mengevaluasi efektivitas model hibrid *Cosine Similarity* dan *Model* (LLM) dalam meningkatkan akurasi dan konsistensi penilaian esai, serta membandingkan hasilnya dengan penilaian manual yang dilakukan oleh penguji manusia.

1.4 Manfaat Penelitian

1. Penelitian ini bermanfaat bagi institusi pendidikan dalam meningkatkan efisiensi proses evaluasi akademik. Melalui sistem penilaian otomatis, beban administratif dosen dapat dikurangi sekaligus meminimalkan potensi bias penilaian. Hal ini mendukung terciptanya proses evaluasi yang lebih adil dan konsisten, serta mengoptimalkan penggunaan sumber daya dalam kegiatan akademik.
2. Manfaat bagi mahasiswa terletak pada ketersediaan umpan balik yang cepat, jelas, dan objektif. Dengan adanya sistem ini, mahasiswa dapat lebih mudah mengidentifikasi kelebihan dan kekurangan dalam jawaban esai mereka. Hasil penilaian yang konsisten juga membantu mahasiswa dalam memahami standar akademik yang berlaku serta mendorong peningkatan kualitas pembelajaran mandiri.
3. Penelitian ini memberikan kontribusi akademik melalui pengembangan metode penilaian esai berbasis pemrosesan bahasa alami. Pendekatan yang menggabungkan *Cosine Similarity* dengan analisis semantik diharapkan dapat membuka peluang penelitian lanjutan, baik dalam bidang evaluasi berbasis teks maupun pengembangan sistem pembelajaran digital. Dengan

demikian, penelitian ini tidak hanya menawarkan solusi praktis, tetapi juga memperkaya literatur akademik di bidang pendidikan.

1.5 Batasan Masalah

1. Penelitian ini fokus pada penilaian esai dalam bahasa Indonesia. Implementasi dalam bahasa lain mungkin memerlukan penyesuaian metode yang berbeda.
2. Kapasitas komputasi terbatas untuk menjalankan *Model* (LLM) dan Kebutuhan sumber daya komputasi tinggi untuk pemrosesan dan inferensi model.
3. Ketersediaan dan kualitas data esai yang digunakan untuk melatih dan menguji sistem penilaian otomatis terbatas pada *dataset* yang tersedia, yang mungkin tidak mencakup semua variasi tipe soal dan karakteristik jawaban mahasiswa.
4. Efektivitas sistem sangat bergantung pada kemampuan dan perkembangan terbaru dari *Model* (LLM). Perkembangan teknologi LLM yang cepat dapat mempengaruhi relevansi metode yang digunakan.
5. Sistem mungkin mengalami kesulitan dalam menilai esai dengan topik yang sangat bervariasi atau tipe soal yang memiliki karakteristik unik, seperti yang diidentifikasi dalam penelitian sebelumnya.
6. Studi ini lebih fokus pada pengembangan dan evaluasi sistem secara teoretis dan eksperimental. Implementasi praktis di lingkungan pendidikan nyata memerlukan pertimbangan tambahan seperti integrasi dengan sistem yang ada dan pelatihan pengguna.