

BAB I

PENDAHULUAN

1.1 Latar Belakang

Dalam proses pembelajaran, evaluasi merupakan komponen penting untuk mengukur pencapaian kompetensi serta kemampuan berpikir peserta didik secara menyeluruh [1]. Salah satu bentuk evaluasi yang banyak digunakan adalah tes esai, karena memungkinkan peserta didik mengungkapkan gagasan dan pemahaman melalui uraian tertulis yang membutuhkan keterampilan berpikir analitis dan sintesis [2]. Tes esai mendorong peserta didik berpikir kritis dan reflektif dalam merespons masalah, sehingga dapat menilai kemampuan berpikir tingkat tinggi secara lebih komprehensif dibandingkan evaluasi objektif semata [3]. Hal ini menunjukkan bahwa tes esai berperan penting dalam mengembangkan keterampilan berpikir tingkat tinggi dan kemampuan menyusun argumen yang relevan dalam konteks pembelajaran [4].

Penilaian esai yang dilakukan secara manual masih menjadi tantangan dalam proses evaluasi pembelajaran karena melibatkan proses pembacaan, pemahaman, dan pemberian skor secara individual untuk setiap jawaban peserta didik. Penilaian manual juga berpotensi menimbulkan subjektivitas antar penilai yang dapat memengaruhi konsistensi dan objektivitas hasil evaluasi [5], [6]. Seiring dengan perkembangan pembelajaran berbasis digital dan *e-learning*, sistem evaluasi dituntut mampu bekerja secara cepat, efisien, serta mengelola data hasil belajar secara otomatis agar proses penilaian dapat dilakukan secara lebih sistematis dan terintegrasi melalui platform teknologi pendidikan berbasis web [7], [8]. Namun dalam praktiknya, banyak institusi pendidikan masih menghadapi keterbatasan dalam penerapan sistem evaluasi berbasis teknologi karena kurangnya pemanfaatan platform digital yang mampu mengotomatisasi proses penilaian dan analisis hasil belajar secara efektif [7]. pengembangan sistem *Automatic Essay Scoring (AES)* berbasis website dapat menjadi salah satu solusi untuk meningkatkan efisiensi penilaian, mengurangi subjektivitas penilai, serta menghasilkan evaluasi yang lebih cepat dan konsisten dalam lingkungan pembelajaran digital [8], [9].

Penelitian mengenai penilaian esai otomatis terus berkembang untuk mengatasi keterbatasan penilaian manual dalam proses evaluasi pembelajaran. Salah satu pendekatan yang banyak digunakan adalah representasi teks berbasis transformer, seperti *Sentence-BERT*, yang mampu mengubah teks menjadi vektor semantik sehingga hubungan makna

antara jawaban pengguna dan kunci jawaban dapat dianalisis menggunakan metode *Cosine Similarity* [10], [11]. *BERT* dan berbagai model turunannya yang berbasis arsitektur transformer terbukti efektif dalam memahami konteks serta makna teks secara dua arah, sehingga menghasilkan representasi teks yang lebih akurat dibandingkan metode tradisional seperti *TF-IDF*, meskipun jawaban ditulis dengan variasi penulisan yang berbeda [10], [11]. Melalui pendekatan ini, tingkat kemiripan antara jawaban esai dan kunci jawaban dapat dihitung secara matematis dan diinterpretasikan sebagai skor penilaian yang lebih objektif dan konsisten [12], [13]. Penerapan metode tersebut dalam sistem penilaian esai berbasis web memungkinkan proses evaluasi dilakukan secara otomatis berdasarkan kemiripan konteks, sehingga dapat mempercepat proses penilaian sekaligus mengurangi beban kerja evaluasi yang sebelumnya dilakukan secara manual [10], [14].

Beberapa penelitian terdahulu telah mengkaji pengembangan sistem *Automatic Essay Scoring* (AES). Chamidah et al. mengembangkan model penilaian esai berbasis SBERT, namun masih berupa model eksperimental tanpa implementasi web [15]. Pradani dan Suadaa menerapkan *Semantic Textual Similarity* berbasis *transformer*, tetapi berhenti pada tahap pengujian model dan belum diwujudkan dalam antarmuka web [16]. Pane et al. membangun sistem penilaian esai menggunakan *Cosine Similarity*, namun belum dikemas dalam platform website yang terintegrasi dengan alur penilaian pengajar [17]. Azzaroh et al. mengembangkan AES berbasis web menggunakan BERT dan *Levenshtein Distance*, tetapi implementasinya lebih menekankan kemiripan karakter dibandingkan kemiripan semantic [9]. Berdasarkan kajian tersebut, terdapat celah penelitian berupa belum tersedianya sistem AES berbasis *website* yang mengintegrasikan *Sentence-BERT* dan *Cosine Similarity* yang lebih menekankan kemiripan makna dalam satu platform digital, sehingga penelitian ini berfokus pada implementasi system.

Berdasarkan permasalahan yang telah dipaparkan pada pembahasan sebelumnya, penelitian ini mengangkat judul **“Implementasi Sistem *Automatic Essay Scoring* Berbasis Website Menggunakan *Sentence-BERT* dan *Cosine Similarity*”** sebagai upaya untuk mengatasi kendala dalam proses penilaian jawaban esai. Sistem yang dikembangkan diharapkan mampu membantu dosen dalam melakukan evaluasi secara lebih cepat, objektif, dan konsisten, sehingga proses penilaian jawaban esai dapat berlangsung lebih efektif tanpa bergantung sepenuhnya pada metode manual.

1.2 Rumusan Masalah

Berdasarkan uraian latar belakang yang telah dipaparkan, permasalahan yang menjadi fokus dalam penelitian ini adalah penilaian esai yang masih manual menyebabkan proses evaluasi kurang efisien dan konsisten, sehingga dibutuhkan sistem otomatis berbasis web yang mampu menganalisis kemiripan makna antarjawaban secara cepat dan akurat.

1.3 Tujuan

Penelitian ini bertujuan untuk:

1. Mengembangkan dan menerapkan sistem *Automatic Essay Scoring* berbasis website dengan memanfaatkan model *Sentence-BERT* dan metode *Cosine Similarity* dalam menilai jawaban esai secara otomatis.
2. Menghasilkan sistem penilaian esai yang mampu merepresentasikan jawaban secara kontekstual serta mengukur tingkat kemiripan antara jawaban peserta didik dan kunci jawaban secara lebih konsisten dan terukur.

1.4 Manfaat

Adapun manfaat yang diharapkan dari penelitian ini dapat diuraikan sebagai berikut:

1. Sistem dapat membantu mempercepat proses penilaian jawaban esai secara otomatis sehingga mengurangi beban kerja evaluasi yang sebelumnya dilakukan secara manual dan meminimalkan subjektivitas dalam pemberian skor.
2. Sistem memberikan hasil penilaian yang lebih konsisten, objektif, dan transparan sehingga mahasiswa dapat memperoleh umpan balik evaluasi secara lebih cepat.
3. Sistem dapat dijadikan alternatif platform evaluasi berbasis digital yang mendukung penerapan *e-learning* secara lebih terintegrasi dan efisien.
4. Penelitian ini memberikan kontribusi referensi dalam pengembangan sistem *Automatic Essay Scoring* berbasis *Natural Language Processing*, khususnya penerapan *Sentence-BERT* dan *Cosine Similarity* untuk Bahasa Indonesia.

1.5 Ruang Lingkup

Berikut adalah ruang lingkup dari penelitian ini:

1. Dataset yang digunakan dalam penelitian ini adalah *Indonesian Query Answering Dataset for Online Essay Test System* [18]. Jumlah soal yang dipakai adalah 10 soal dan 10 referensi jawaban yang sudah ada di dataset. total data yang dipakai adalah 515 jawaban siswa.

2. Sistem hanya memproses input berupa teks naratif yang diisikan dalam kolom input teks minimal 10 karakter.
3. Sistem ini hanya diterapkan pada mata pelajaran yang menghasilkan jawaban berupa teks naratif atau esai, seperti Bahasa Indonesia, PPKn, Sosiologi, dan Sejarah.
4. Pengembangan sistem menggunakan bahasa pemrograman *Python* dengan lingkungan pengembangan *Google Colaboratory* untuk proses pelatihan model dan *Flask* sebagai *framework backend* berbasis web
5. Basis data yang digunakan adalah *SQLite* untuk penyimpanan data pengguna dan hasil penilaian.
6. *Epoch* yang digunakan dalam penelitian ini berada pada rentang 2 hingga 5 epoch untuk menjaga keseimbangan performa model pada proses fine-tuning, dengan nilai epoch 1 ditetapkan konstan pada tahap pencarian *learning rate* optimal.
7. *Learning rate* yang diuji dibatasi pada rentang 1×10^{-6} hingga 5×10^{-4} . Enam nilai awal (1×10^{-6} , 2×10^{-6} , 5×10^{-6} , 1×10^{-5} , 2×10^{-5} , 5×10^{-5}) mengacu pada rentang yang digunakan dalam penelitian rujukan, ditambah tiga nilai variasi (1×10^{-4} , 2×10^{-4} , 5×10^{-4}) untuk memastikan titik optimal *learning rate* teruji secara menyeluruh [19].
8. *Batch size* yang digunakan adalah 16 untuk memastikan efisiensi penggunaan memori sekaligus menjaga kestabilan proses pembaruan bobot model.
9. Output sistem berupa nilai *cosine similarity* dalam rentang 0–1 yang dikonversi menjadi skor akhir dalam skala 0–100 menggunakan rumus $\text{Skor Akhir} = \text{Cosine Similarity} \times 100$
10. Tingkat kemiripan antarvektor dihitung menggunakan metode *Cosine Similarity*.
11. Sistem menampilkan kategori interpretasi hasil penilaian berdasarkan rentang skor, yaitu 0–20 (Sangat Kurang), 21–40 (Kurang), 41–60 (Cukup), 61–80 (Baik), dan 81–100 (Sangat Baik) [48].
12. Pengujian sistem dilakukan dengan membandingkan skor yang dihasilkan sistem dengan skor referensi menggunakan metrik evaluasi *Root Mean Square Error* (RMSE).