E-ISSN: 3025-0994
Halaman 5050-5059
Volume 4 Nomor 4 Tahun 2026

Penerapan Algoritma Sentence-BERT dan Cosine Similarity pada Sistem Penilaian Otomatis Jawaban Esai Terbatas

Alvianto Ramadana[1], Jumadil Nangi[2], La Surimi[3], Ilham Julian Efendi[4]

[1,2,3,4] Jurusan Teknik Informatika, Universitas Halu Oleo Kendari
e-mail: alviantoramadana_k1a123048@student.uho.ac.id, jumadilnangi@uho.ac.id, lasurimi@uho.ac.id, ilham.julian.efendi@uho.ac.id

Abstrak

Penilaian jawaban esai terbatas secara manual memerlukan waktu pemeriksaan yang relatif panjang dan berpotensi menghasilkan perbedaan skor antarpemeriksa. Penelitian ini bertujuan mengembangkan sistem penilaian otomatis berbasis web dengan mengintegrasikan Sentence-BERT dan Cosine Similarity, serta mengevaluasi kinerja model dan penerimaan pengguna. Pengembangan perangkat lunak mengikuti tahapan Rational Unified Process, yaitu inception, elaboration, construction, dan transition. Jawaban siswa dan kunci jawaban diolah menjadi representasi semantik melalui model Sentence-BERT dan mean pooling, kemudian dibandingkan menggunakan Cosine Similarity dan dikonversi menjadi skor. Implementasi aplikasi memanfaatkan Python, Flask, dan SQLite. Evaluasi algoritma dilakukan terhadap 48 jawaban esai siswa program keahlian Teknik Jaringan Komputer dan Telekomunikasi di SMKN 1 Kendari. Kinerja sistem diukur menggunakan akurasi, F1-score, dan Quadratic Weighted Kappa, sedangkan kelayakan penggunaan dinilai melalui User Acceptance Testing terhadap 18 siswa dan seorang guru. Hasil pengujian jawaban siswa nyata memperoleh akurasi sebesar 81,25%, Weighted-F1 sebesar 82,64%, dan Quadratic Weighted Kappa sebesar 0,8125. Tingkat penerimaan pengguna mencapai 88,74%, dengan persentase penerimaan siswa sebesar 89,17% dan guru sebesar 80,00%. Temuan menunjukkan bahwa pendekatan kesamaan semantik dapat mendukung pemeriksaan esai pendek, tetapi validasi pengajar tetap diperlukan, khususnya untuk jawaban ambigu dan variasi konsep yang tidak tercakup dalam kunci jawaban tunggal. Evaluasi ini memberikan bukti awal mengenai pemanfaatan teknologi penilaian berbasis semantik dalam pembelajaran kejuruan.

Kata Kunci: Penilaian Esai Otomatis, Sentence-BERT, Cosine Similarity, Kemiripan Semantik, Website

Abstract

Manual grading of short-answer essay questions is relatively time-consuming and may result in scoring inconsistencies among graders. This study aims to develop a web-based automated grading system by integrating *Sentence-BERT* and *Cosine Similarity*, as well

as to evaluate the model's performance and user acceptance. The software development process follows the *Rational Unified Process* (RUP), which consists of four phases: *inception*, *elaboration*, *construction*, and *transition*. Students' answers and reference answers are processed into semantic representations using the *Sentence-BERT* model and *mean pooling*. The resulting representations are then compared using *Cosine Similarity* and converted into numerical scores. The application is implemented using Python, Flask, and SQLite. The algorithm was evaluated using 48 actual student essay responses from the Computer and Telecommunications Network Engineering program at SMKN 1 Kendari. System performance was measured using accuracy, F1-score, and *Quadratic Weighted Kappa*, while usability was assessed through *User Acceptance Testing* involving 18 students and one teacher. The evaluation of actual student responses yielded an accuracy of 81.25%, a *Weighted-F1* score of 82.64%, and a *Quadratic Weighted Kappa* of 0.8125. Overall user acceptance reached 88.74%, with an acceptance rate of 89.17% among students and 80.00% among teachers. The findings indicate that a semantic similarity approach can support the grading of short-answer essays. However, teacher validation remains necessary, particularly when evaluating ambiguous responses and variations in conceptual explanations that are not covered by a single reference answer. This evaluation provides preliminary evidence of the potential application of semantic-based automated grading technology in vocational education.

Keywords : Automated Essay Scoring, Sentence-BERT, Cosine Similarity, Semantic Similarity, Website

LATAR BELAKANG

Penilaian esai terbatas digunakan untuk mengukur penguasaan konsep melalui jawaban yang disusun dengan bahasa peserta didik sendiri (Sucipta dkk., 2023). Bentuk evaluasi ini memungkinkan pemeriksaan keterkaitan gagasan yang sulit diukur melalui pilihan jawaban tertutup. Namun, keberagaman struktur kalimat dan pilihan kata membuat pemeriksaan membutuhkan waktu serta konsistensi penilaian, terutama ketika jumlah jawaban yang dinilai cukup banyak (Kinanti & Qoiriah, 2020). Kondisi ini menunjukkan bahwa penilaian esai secara manual masih menghadapi kendala efisiensi dan konsistensi (Rokhman dkk., 2025).

Pengembangan sistem penilaian esai otomatis (Automated Essay Scoring/AES) menjadi salah satu solusi untuk mempercepat dan meningkatkan konsistensi penilaian, mengurangi bias subjektif, serta menghemat waktu koreksi pendidik (Hutasuhut dkk., 2025). Namun, pencocokan kata belum mampu mengenali kesamaan makna secara optimal ketika jawaban menggunakan susunan kata berbeda. Kini AES memanfaatkan Natural Language Processing (NLP) untuk menilai kesesuaian makna jawaban dengan kunci jawaban referensi (Tarumingkeng, 2024). Chamidah dkk. (2022) mengkaji penilaian esai menggunakan kesamaan semantik berbasis Sentence-BERT (Chamidah dkk., 2022), sedangkan Azzaroh dkk. (2025) mengembangkan sistem penilaian esai berbasis web menggunakan BERT dan Levenshtein Distance dengan pendekatan semantik-sintaksis (Azzaroh dkk., 2025)..

Perkembangan deep learning, khususnya arsitektur Transformer, meningkatkan kemampuan komputer dalam memahami bahasa alami (Sihombing dkk., 2026). Salah satu model yang berkembang adalah BERT (Bidirectional Encoder Representations from Transformers), yang mampu memahami konteks teks secara dua arah dan menunjukkan kinerja baik dalam berbagai tugas NLP (Zhang & Shafiq, 2024; Gardazi dkk., 2025). Namun, BERT belum dirancang secara khusus untuk mengukur kesamaan antarkalimat. Oleh karena itu, dikembangkan Sentence-BERT (SBERT) yang menghasilkan representasi vektor kalimat untuk mengukur kesamaan semantik secara lebih efektif (Han & Yang, 2024).

SBERT adalah modifikasi arsitektur BERT dengan pendekatan siamese network yang dirancang untuk menghasilkan representasi embedding kalimat yang lebih stabil dan efisien dalam pengukuran kesamaan semantik (Agiharta dkk., 2024). Pendekatan ini memungkinkan perhitungan kesamaan antar jawaban dilakukan secara lebih akurat dan efisien, sehingga lebih sesuai untuk diterapkan dalam sistem penilaian esai otomatis untuk mengukur kesamaan semantik antara jawaban peserta didik dan kunci jawaban (Sitorus, 2026). Pengukuran tersebut dapat dikombinasikan dengan cosine similarity, yaitu ukuran kesamaan arah dua vektor, sebagai dasar penentuan skor (Fathuddin dkk. 2025).

Berdasarkan permasalahan tersebut, penelitian ini mengembangkan aplikasi penilaian otomatis jawaban esai terbatas berbasis web dengan mengintegrasikan SBERT dan cosine similarity. Tujuan penelitian adalah mengevaluasi kesesuaian hasil penskoran otomatis terhadap penilaian guru serta menilai tingkat penerimaan sistem oleh pengguna. Kontribusi penelitian terletak pada integrasi metode kesamaan semantik dalam alur ujian yang dapat digunakan pengajar dan siswa, disertai pengujian kinerja algoritma dan penerimaan pengguna pada konteks pembelajaran kejuruan. Penelitian dibatasi pada jawaban pendek dengan satu kunci referensi, tanpa analisis tata bahasa atau gaya penulisan.

METODE

a. Jenis dan Sumber Data

Data penelitian terdiri atas dataset sekunder dan data operasional sistem. Dataset sekunder berisi pasangan jawaban esai yang memuat pertanyaan, kunci jawaban, jawaban peserta didik, dan skor dengan tiga kategori, yaitu salah, sebagian benar, dan benar. Dataset tersebut di peroleh dari platform github milik Aggrawal dkk yang berisi jawaban singkat yang dikumpulkan dari ujian teknik nyata yang diselenggarakan di sebuah lembaga akademik ternama (Aggarwal dkk., 2025). Data tersebut digunakan untuk melatih atau menyesuaikan parameter SBERT, mengevaluasi model, memilih model terbaik dan menentukan threshold.

Sementara itu, data operasional sistem dikelola melalui basis data relasional SQLite. Data ini mencakup data autentikasi pengguna untuk pengajar dan admin, data konfigurasi ujian, data bank soal dan kunci referensi, data sesi pengerjaan peserta ujian, data teks jawaban esai peserta, serta data hasil kalkulasi penilaian otomatis yang memuat skor kemiripan semantik, label klasifikasi, dan nilai akhir terkonversi dalam skala 0 hingga 100

b. Metode Pengembangan Sistem

Metode pengembangan sistem yang digunakan yaitu Rational Unified Process (RUP). RUP merupakan kerangka kerja proses rekayasa perangkat lunak yang bersifat iteratif dan berorientasi objek yang dirancang untuk menyediakan pendekatan sistematis dalam pengembangan sistem perangkat lunak (Irawan dkk., 2024). Proses-proses dalam model RUP ini meliputi empat fase, yaitu inception untuk identifikasi masalah dan kebutuhan pengguna, elaboration untuk perancangan alur serta pemodelan sistem, construction untuk implementasi, dan transition untuk pengujian serta evaluasi penggunaan (Sintawati, 2022).

c. Implementasi dan Analisis

Tahap implementasi sistem merupakan proses penerapan hasil analisis dan perancangan ke dalam bentuk aplikasi berbasis website. Sistem penilaian esai otomatis dikembangkan dengan menggunakan bahasa pemrograman Python yang didukung oleh framework Flask dan basis data SQLite (Kristianto & Setiyawati, 2021). Pengembangan dilakukan menggunakan Antigravity IDE digunakan sebagai lingkungan pengembangan kode untuk mempermudah proses penulisan, pengelolaan, serta pengembangan kode program secara lebih terstruktur dan efisien.

Pada tahap ini, sistem mengimplementasikan algoritma Sentence-BERT dan Cosine Similarity sebagai inti dari proses penilaian esai otomatis. Dalam penelitian ini, Hugging Face dimanfaatkan sebagai platform utama untuk mengakses dan mengimplementasikan model SBERT yang akan digunakan untuk mengubah teks jawaban siswa dan kunci jawaban referensi menjadi vektor embedding semantik berdimensi padat (Moreno, 2021). Sedangkan metode Cosine Similarity diterapkan untuk menghitung tingkat kemiripan makna antara kedua representasi vektor tersebut (Iskandar & Kurniawati, 2025). Mekanisme dynamic thresholding diterapkan untuk mengklasifikasikan skor kemiripan ke dalam kategori nilai yang objektif dan konsisten.

Pada tahap pemodelan Artificial Intelligence (AI), penelitian ini menggunakan dataset pasangan jawaban esai yang sudah didapatkan. Setiap entri data terdiri atas empat atribut utama, yaitu pertanyaan (question), kunci jawaban referensi yang disusun oleh pendidik (reference_answer), jawaban esai peserta didik (student_answer), serta skor penilaian acuan (score) berskala ordinal yang merepresentasikan tiga kategori penilaian manusia, yaitu skor 0 untuk jawaban salah, skor 1 untuk jawaban sebagian benar, dan skor 2 untuk jawaban benar. Sebelum digunakan, dataset melewati serangkaian tahap prapemrosesan data (data preprocessing), meliputi penghapusan data bertipe null, normalisasi teks (konversi huruf kecil dan eliminasi spasi berlebih), serta pembersihan data duplikat, untuk menghasilkan data bersih dan valid.

Untuk menjaga integritas evaluasi dan mencegah terjadinya kebocoran informasi (data leakage), pembagian data dilakukan menggunakan metode Group Shuffle Split dengan menetapkan gabungan antara pertanyaan dan kunci jawaban sebagai kunci kelompok (group key). Melalui pendekatan ini, seluruh variasi jawaban peserta untuk soal tertentu hanya akan berada di salah satu bagian data (eksklusif), sehingga model dievaluasi menggunakan butir soal yang belum pernah ditemui pada fase pelatihan. Dataset tersebut dibagi ke dalam tiga subset data, yaitu data latih (training set) sebanyak 71% untuk proses

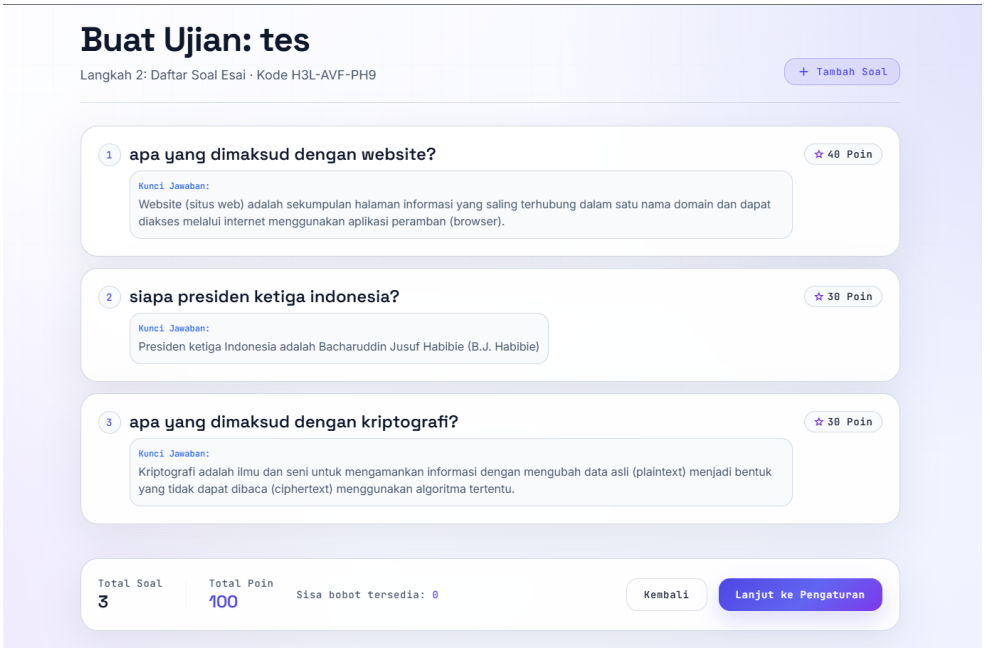
fine-tuning model Sentence-BERT menggunakan fungsi kerugian Cosine Similarity Loss, data validasi (validation set) sebanyak 12% untuk mengoptimalkan ambang batas klasifikasi (dynamic threshold tuning), dan data uji (testing set) sebanyak 17% untuk mengukur keandalan dan generalisasi model.

Model dasar (pre-trained backbone) yang digunakan adalah paraphrase-multilingual-mpnet-base-v2, yaitu model Sentence-BERT multibahasa berbasis Transformer yang memetakan kalimat ke dalam ruang vektor berdimensi 768. Pasangan teks dienkode menggunakan skema asymmetric context encoding dengan menggabungkan teks pertanyaan pada kalimat kunci dan jawaban.

HASIL DAN PEMBAHASAN

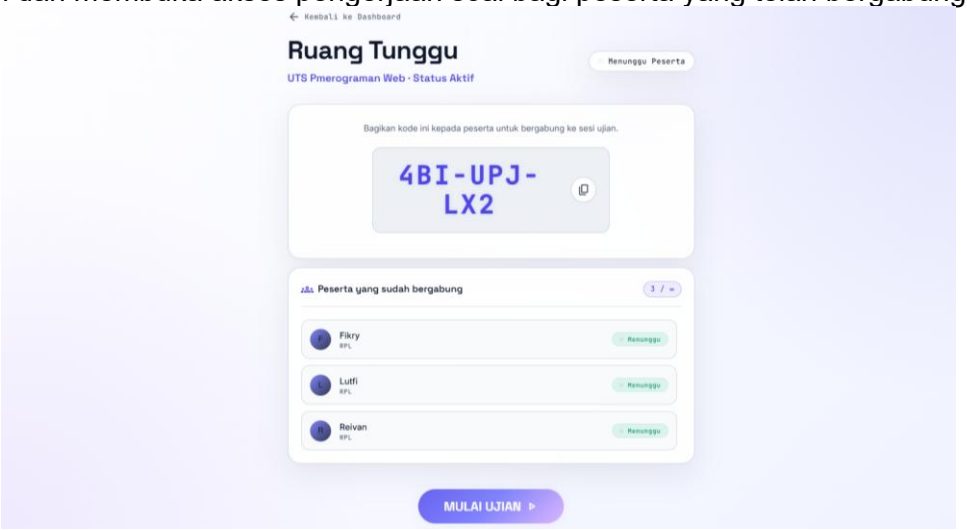
Implementasi menghasilkan aplikasi yang menghubungkan proses penyelenggaraan ujian dengan penilaian berbasis kesamaan semantik. Pengajar dapat membuat soal dan kunci jawaban, memberikan kode ujian, memantau pelaksanaan, dan mengakses rekapitulasi skor. Jawaban yang dikirimkan peserta diproses oleh modul SBERT dan cosine similarity, kemudian hasilnya disimpan serta ditampilkan kembali pada aplikasi.

Gambar 1 menunjukkan tampilan pengelolaan soal dan jawaban referensi yang digunakan sebagai masukan penilaian. Penggunaan kunci tunggal memudahkan pengolahan otomatis; pada saat yang sama, desain tersebut membatasi variasi formulasi jawaban benar yang dapat dijadikan pembanding.



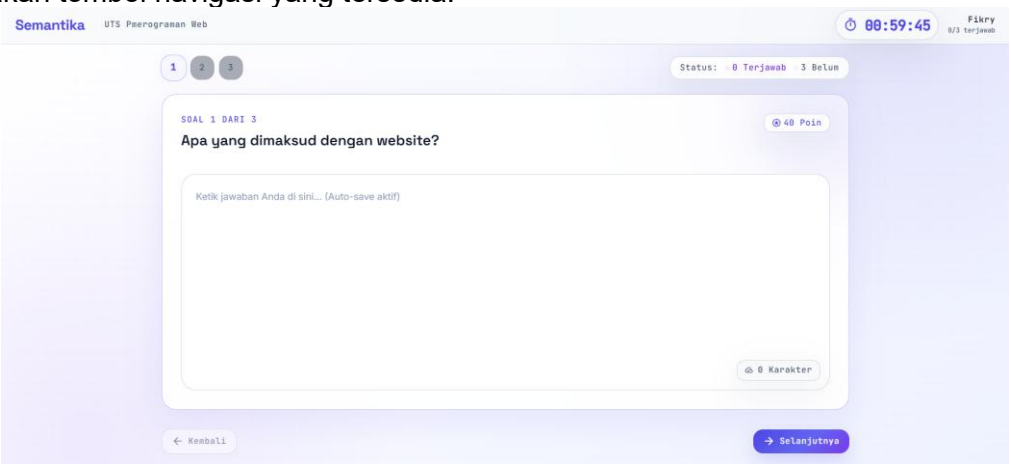
Gambar 1. Halaman Kelola Soal

Selanjutnya, halaman ruang tunggu pada Gambar 2 digunakan pengajar untuk memantau kesiapan peserta sebelum ujian dimulai. Halaman ini menampilkan judul dan status ujian, kode akses yang dapat disalin, serta daftar peserta yang telah bergabung beserta informasi kelas dan status kesiapannya. Tombol Mulai Ujian digunakan untuk memulai sesi dan membuka akses pengerjaan soal bagi peserta yang telah bergabung.



Gambar 2. Halaman Ruang Tunggu

Setelah ujian dimulai, peserta mengerjakan soal melalui halaman jawab soal sebagaimana ditampilkan pada Gambar 3. Antarmuka ini menyediakan informasi judul ujian, identitas peserta, waktu pengerjaan, navigasi nomor soal, dan indikator progres. Peserta dapat membaca pertanyaan, memasukkan jawaban melalui kolom teks yang dilengkapi fitur penyimpanan otomatis (auto-save) dan penghitung karakter, serta berpindah antarsoal menggunakan tombol navigasi yang tersedia.



Gambar 2. Halaman Jawab Soal

Hasil pengujian kinerja algoritma dirangkum pada Tabel 1. Pengujian dilakukan terhadap 48 jawaban siswa pada ujian nyata yang telah divalidasi oleh guru. Evaluasi kinerja algoritma menggunakan sejumlah indikator untuk mengetahui tingkat kesesuaian hasil penilaian otomatis dengan penilaian guru. Hasil pengujian tersebut digunakan untuk menggambarkan kemampuan algoritma dalam melakukan penilaian jawaban siswa berdasarkan tingkat kesamaan semantik.

Tabel 1. Ringkasan hasil pengujian kinerja algoritma

Metrik	Ujian (n = 48)
Akurasi	81,25%
F1-score	Weighted-F1 82,64%
Quadratic Weighted Kappa	0,8125

Tabel 1 memperlihatkan nilai QWK sebesar 0,8125 pada pengujian menggunakan data ujian nyata. Dalam konteks penelitian ini, QWK digunakan untuk menunjukkan tingkat kesepakatan antara skor yang dihasilkan sistem dan skor yang diberikan oleh penilai, bukan sebagai ukuran akurasi. Capaian tersebut mendukung penggunaan SBERT dan cosine similarity sebagai dasar penskoran otomatis. Namun, nilai QWK tersebut belum menunjukkan bahwa setiap kategori penilaian memiliki kinerja yang sama baiknya. Oleh karena itu, hasil F1-score untuk setiap kategori disajikan secara lebih rinci pada Tabel 2.

Tabel 2. F1-score dan jumlah sampel menurut kategori jawaban

Kategori	F1	Sampel
Salah	0,5000	4
Sebagian Benar	0,4706	9
Benar	0,9552	35

Pada pengujian lapangan, kategori Benar memperoleh F1-score sebesar 0,9552 dengan 35 sampel, sedangkan kategori Salah dan Sebagian Benar masing-masing memperoleh F1-score sebesar 0,5000 dan 0,4706 dengan jumlah sampel yang lebih sedikit. Hasil ini menunjukkan bahwa sistem memiliki kinerja yang sangat baik dalam mengidentifikasi jawaban yang sepenuhnya benar, tetapi masih menghadapi kesulitan dalam membedakan jawaban yang salah dan jawaban yang hanya memenuhi sebagian konsep. Selain itu, nilai akurasi sebesar 81,25% perlu diinterpretasikan dengan mempertimbangkan distribusi kategori jawaban, yaitu 35 dari 48 jawaban termasuk kategori Benar. Dominasi kategori tersebut dapat memengaruhi tingginya nilai akurasi secara keseluruhan. Dengan demikian, capaian akurasi yang tinggi belum sepenuhnya mencerminkan kemampuan sistem dalam mengenali seluruh kategori jawaban secara seimbang. Pengujian lapangan ini melibatkan 16 siswa dengan tiga butir soal, sehingga hasilnya masih terbatas pada konteks

pengujian tersebut dan belum dapat digeneralisasi untuk seluruh mata pelajaran maupun berbagai bentuk pertanyaan.

Pengujian pada lingkungan sekolah juga menghasilkan jenis bukti yang berbeda dari pengujian algoritma murni. Ketika siswa mengerjakan ujian, kinerja penggunaan sistem mencakup kejelasan navigasi, kemudahan memasukkan jawaban, dan ketersediaan rekap hasil bagi guru. Namun, indikator UAT dalam penelitian ini mengukur persepsi responden setelah mencoba aplikasi, bukan kecepatan pemeriksaan berdasarkan pengukuran waktu sebelum dan sesudah penggunaan. Dengan membedakan kedua hal tersebut, simpulan kelayakan dapat tetap sesuai bukti yang tersedia dan tidak berkembang menjadi klaim efisiensi kuantitatif yang belum diuji secara khusus.

Tabel 3. Hasil User Acceptance Testing sistem Semantika

Responden	n	Skor	Maksimum	Persentase
Siswa	18	642	720	89,17%
Guru	1	28	35	80,00%
Gabungan	19	670	755	88,74%

Berdasarkan Tabel 3, penerimaan siswa sebesar 89,17% berada pada kategori Sangat Layak, sedangkan persentase guru 80,00% berada pada kategori Layak. Persentase gabungan 88,74% diperoleh dari 670 dibagi 755, sehingga nilainya mencerminkan bobot jumlah butir dan responden pada kedua kelompok. Temuan ini menunjukkan tanggapan positif atas fungsi aplikasi yang dicoba responden, bukan bukti bahwa semua skor otomatis telah benar atau bahwa sistem pasti menghemat waktu koreksi pada seluruh kondisi.

SIMPULAN

Sistem Semantika berhasil mengintegrasikan Sentence-BERT dan cosine similarity untuk penilaian otomatis jawaban esai terbatas berbasis web. Pengujian terhadap 48 jawaban nyata memperoleh akurasi sebesar 81,25% dan QWK sebesar 0,8125. Sementara itu, hasil User Acceptance Testing (UAT) menunjukkan tingkat penerimaan gabungan sebesar 88,74%. Hasil tersebut menunjukkan bahwa sistem berpotensi membantu proses pemeriksaan awal jawaban esai secara otomatis. Meskipun demikian, hasil penilaian otomatis tetap memerlukan validasi pengajar, khususnya dalam mengidentifikasi jawaban yang hanya memenuhi sebagian konsep. Penelitian selanjutnya disarankan untuk memperluas jumlah sampel, menggunakan beberapa referensi jawaban, serta mengevaluasi keterpenuhan konsep dan konsistensi skor pada beragam bentuk soal.

DAFTAR PUSTAKA

Aggarwal, D., Sil, P., Raman, B. & Bhattacharyya, P. 2025. “I Understand Why I Got This Grade”: Automatic Short Answer Grading (ASAG) with Feedback. International Conference on Artificial Intelligence in Education, 15879: 304–318.

Agiharta, K.F., Suteja, B.R. & Ayub, M. 2024. Penerapan Sentence-BERT untuk Similaritas Kompetensi Pekerjaan dan Mata Kuliah. Jurnal Teknik Informatika dan Sistem Informasi, 10(3): 449–467.

Chamidah, N., Santoni, M.M., Irmanda, H.N. & Astriratma, R. 2022. Penilaian Esai Pendek Otomatis Berdasarkan Kemiripan Semantik dengan SBERT. Tekno.Com, 21(4): 732.

Fathuddin, M.A.H., Mandyartha, E.P. & Nurlaili, A.L. 2025. Penerapan Kalimat-BERT dan Cosine Kemiripan untuk Pencarian Semantik Dokumen Skripsi dalam Format PDF. Ranah Research: Journal of Multidisciplinary Research and Development, 8(1): 322–337.

Febrianti Azzaroh, A., Surya Editya, A. & Khoir Al-Haq, A. 2025. Sistem Penilaian Esai Otomatis Berbasis Web Menggunakan Model BERT dan Levenshtein Distance. Jurnal Pustaka AI (Pusat Akses Kajian Teknologi Artificial Intelligence), 5(2): 142–148.

Gardazi, N.M. dkk. 2025. BERT Applications in Natural Language Processing: A Review. Artificial Intelligence Review, 58(6): 1–49.

Han, J. & Yang, L. 2024. Sentence Embedding Generation Framework Based on Kullback–Leibler Divergence Optimization and RoBERTa Knowledge Distillation. Mathematics, 12(24): 1–20.

Hutasuhut, S. dkk. 2025. Pemanfaatan Perkembangan Teknologi dalam Penilaian Otomatis untuk Meningkatkan Efisiensi Kerja Guru di MTsS Ikaba Paluh Manis Gebang. Pema, 5(1): 669–676.

Irawan Chandra, Y., Irawati, D.R. & Riastuti, M. 2024. Penerapan Model Rational Unified Process (RUP) dalam Membangun Aplikasi Promosi dan Penjualan Berbasis Web (Studi Kasus: Toko Mie Ayam “Neng Kene Kiyeh”). Ikraith Informatika, 8(3): 107–120.

Iskandar, D. & Kurniawati, A. 2025. Analisis Perbandingan Teknik Word2Vec dan Doc2Vec dalam Mengukur Kemiripan Dokumen Menggunakan Cosine Similarity. Jurnal Teknologi Informasi dan Ilmu Komputer, 12(1): 133–144.

Kinanti, N.L. & Qoiriah, A. 2020. Sistem Penilaian Otomatis Jawaban Esai Bahasa Indonesia Berdasarkan Kemiripan Kalimat Menggunakan Syntactic-Semantic Similarity. Journal of Informatics and Computer Science, 2(2): 136–143.

Kristianto, E.H.Y. & Setiyawati, N. 2021. Pembangunan Aplikasi Virtual Inventory System (VIS) Berbasis Web Menggunakan Flask Framework (Studi Kasus: PT XYZ). Jurnal Mnemonic, 4(2): 128–135.

Moreno, B.S. 2021. Transformers BERT para Question-Answering sobre COVID-19. Tesis. Lenguajes y Sistemas Informáticos, Universidad Nacional de Educación a Distancia, Madrid.

Sintawati, I.D. 2022. Komparasi Metode RAD dengan RUP pada Pengembangan Sistem Informasi. AKRAB JUARA, 7(2): 94–100.

Sitorus, T.M., Ramadhan, E. & Kasyidi, F. 2026. Penilaian Otomatis Jawaban Esai SMA Menggunakan Sentence-BERT dan Hybrid Levenshtein-Jaccard dengan Akurasi Hybrid. Jurnal Algoritma, 23(1): 39–50.

Sihombing, M. A., Sinaga, I. L., Hutagaol, O. K., Simamora, J. T. J., & Sihite, A. S. 2026. Tinjauan Arsitektur Transformer Dan Penerapannya Pada Pemrosesan Bahasa Alami. IKRA-ITH Informatika: Jurnal Komputer dan Informatika, 10(2), 652-661.

Sucipta, I.W., Candiasa, I.M. & Sudirtha, I.G. 2023. Pengaruh Model Pembelajaran Berbasis Masalah dan Bentuk Asesmen Formatif terhadap Kemampuan Berpikir Kritis. Jurnal Penelitian dan Evaluasi Pendidikan Indonesia, 13(2): 168–178.

Rokhman, N., Maulan, P. A., & Wirahuda, N. A. 2025. Analisis penilaian esai secara otomatis menggunakan Natural Language Processing (NLP) dan Cosine Similarity. Go Infotech: Jurnal Ilmiah STMIK AUB, 31(1), 41-52.

Tarumingkeng, R. C. 2024. Natural Language Processing (NLP). RUDYCT e-PRESS, no.

Zhang, H. & Shafiq, M.O. 2024. Survey of Transformers and Towards Ensemble Learning Using Transformers for Natural Language Processing. Journal of Big Data, 11(1): 1–45.