

Innovation in Automated Essay Scoring for Writing Assessment: A Case Study at SMA Negeri 1 Genteng

Inovasi dalam Penilaian Menulis melalui Skor Esai Otomatis: Studi Kasus di SMA Negeri 1 Genteng

Lestari Andhini*¹, Reza Maulana ², Naufal Rizky Pratama ³

^{1,2,3} Universitas Negeri Malang, Indonesia

Received: 22-08-2025 | Revised: 12-09-2025 | Accepted: 12-10-2025

Abstrak

Penelitian ini bertujuan mengembangkan evaluasi sistem Automated Essay Scoring (AES) berbasis Bidirectional Long Short-Term Memory (Bi-LSTM) untuk meningkatkan efisiensi dan objektivitas penilaian tulisan siswa. Studi kasus dilakukan di SMA Negeri 1 Genteng dengan menggunakan data esai siswa yang dinilai oleh guru sebagai ground truth. Hasil evaluasi menunjukkan nilai QWK 0,83, MAE 0,41, dan PCC 0,86, yang menunjukkan kesesuaian tinggi antara skor model dan penilaian manual. Selain hasil kuantitatif, persepsi guru dan siswa menunjukkan bahwa sistem AES dinilai efisien, objektif, dan membantu proses pembelajaran menulis. Sebanyak 100% guru menyatakan sistem mempercepat penilaian, sementara 90% siswa menilai umpan balik otomatis membantu perbaikan tulisan. Namun, aspek kreativitas masih memerlukan penilaian manusia. Secara keseluruhan, AES berbasis Bi-LSTM layak digunakan sebagai alat bantu penilaian cerdas di lingkungan pendidikan menengah.

Kata Kunci: Penilaian Esai Otomatis; Bi-LSTM; Persepsi Guru; Penilaian Menulis; Kecerdasan Buatan

This study aims to develop to evaluate a Bidirectional Long Short-Term Memory (Bi-LSTM) based Automated Essay Scoring (AES) system to enhance the efficiency and objectivity of student writing assessment. A case study was conducted at SMA Negeri 1 Genteng using student essays graded by teachers as the ground truth. Evaluation results show QWK 0.83, MAE 0.41, and PCC 0.86, indicating a high agreement between model and manual scores. Beyond quantitative results, teachers and students perceived AES as efficient, objective, and supportive of writing improvement. All teachers (100%) agreed that AES speeds up grading, and 90% of students stated that automatic feedback helps them revise their writing. However, creativity still requires human judgment. Overall, the Bi-LSTM based AES system is feasible as an intelligent assessment tool in secondary education.

Keywords: Automated Essay; Scoring, Bi-LSTM; Teacher Perception; Writing Assessment; Artificial Intelligence

Corresponding Author: andhini.lestari@gmail.com

How to Cite:

Andhini, L., Maulana, R., & Pratama, N.R. (2025). Innovation in Automated Essay Scoring for Writing Assessment: A Case Study at SMA Negeri 1 Genteng. *JINEA: Journal of Innovation in Education and Learning*, 1(3), 143-152. <https://doi.org/10.66031/jinea.v1i3.35>

Copyright ©2025 to the Author. Published by CV. Ihsan Cahaya Pustaka
This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license



1. PENDAHULUAN

Kemampuan menulis merupakan salah satu keterampilan utama dalam pembelajaran bahasa yang berperan penting dalam membentuk kemampuan berpikir kritis, kreativitas, dan komunikasi siswa (Aini et al., 2025; Fatqurhohman et al., 2025). Namun, dalam konteks pendidikan di sekolah menengah, penilaian kemampuan menulis sering kali menjadi tantangan tersendiri karena memerlukan waktu yang lama, tenaga yang besar, serta adanya kemungkinan subjektivitas dalam penilaian (Margareta et al., 2025; Semba, 2024). Selain itu, variasi interpretasi terhadap rubrik penilaian dan beban administrasi guru memperburuk akurasi serta konsistensi hasil evaluasi tulisan siswa (Chamidah et al., 2023; Hiver et al., 2024). Kondisi tersebut menunjukkan perlunya pendekatan yang lebih efisien, objektif, dan inovatif dalam proses penilaian menulis di sekolah.

Kemajuan pesat dalam bidang *Artificial Intelligence (AI)* dan *Natural Language Processing (NLP)* telah mendorong pengembangan sistem *Automated Essay Scoring (AES)* sebagai inovasi dalam penilaian kemampuan menulis (Hussein et al., 2019; Jayadianti et al., 2023). *AES* merupakan sistem berbasis komputer yang memanfaatkan model pembelajaran mesin untuk memberikan skor terhadap teks esai berdasarkan aspek linguistik, semantik, dan struktur tulisan (Chamidah et al., 2023; Pradani & Suadaa, 2023; Verdikha et al., 2024). Di tingkat global, penggunaan *AES* telah terbukti membantu mempercepat proses penilaian sekaligus menjaga konsistensi antarpemilai (Hussein et al., 2019; Susilawati et al., 2022; Wiratmo & Fatichah, 2020). Dalam penelitian oleh Pradani & Suadaa (2023) yaitu menghitung kemiripan semantik teks berbahasa Indonesia antara jawaban esai dan kunci jawabannya yaitu model berbasis *Transformers IndoBERT* potensi kuat penerapan *AES* sebagai inovasi asesmen otomatis yang relevan bagi konteks pendidikan Indonesia.

Penelitian-penelitian terkini semakin memperkuat efektivitas pendekatan tersebut. Rajagede (2021) dalam penelitiannya menemukan bahwa kombinasi berbagai representasi mampu meningkatkan akurasi evaluasi esai dibandingkan penggunaan satu model saja. Hasil serupa diperoleh oleh Pradani & Suadaa (2023), yang menunjukkan bahwa model *Transformer* pra-latih seperti *IndoBERT* memberikan hasil penilaian yang lebih konsisten dan efisien dibandingkan penilaian manual guru. Sementara itu, Semba (2024) menambahkan bahwa model *LSTM* dapat digunakan tidak hanya untuk penilaian skor tetapi juga untuk menghasilkan umpan balik otomatis berbahasa Indonesia yang relevan dengan isi tulisan siswa. Hal ini sejalan dengan penelitian Pradani & Suadaa (2023) yang menegaskan bahwa pendekatan semantic textual similarity berbasis *Transformer* dapat meningkatkan reliabilitas sistem penilaian otomatis untuk tugas esai Bahasa Indonesia.

Walaupun secara teknis telah terbukti efektif, penerapan *AES* dalam konteks sekolah menengah di Indonesia masih sangat terbatas. Hambatan utamanya bukan hanya pada ketersediaan teknologi, tetapi juga pada kesiapan institusi dan penerimaan pengguna, terutama guru dan siswa, terhadap sistem penilaian berbasis AI (Dhini et al., 2023; Rajagede, 2021). Selain itu, hasil penelitian oleh (Pradani & Suadaa, 2023; Ramesh & Sanampudi, 2022; Verdikha et al., 2024) menunjukkan bahwa kualitas tokenisasi kalimat dan data latihan turut memengaruhi keakuratan sistem *AES*, sehingga integrasinya di ruang kelas perlu disertai dengan adaptasi pedagogis yang tepat. Oleh karena itu, inovasi pendidikan yang

menggabungkan penilaian otomatis dan pendekatan pembelajaran menulis manusiawi perlu dikaji lebih jauh agar penerapannya benar-benar efektif secara akademik maupun praktis.

Berdasarkan uraian di atas, penelitian ini berfokus pada inovasi dalam penerapan sistem penilaian esai otomatis (*AES*) untuk mendukung penilaian kemampuan menulis siswa di SMA Negeri 1 Genteng. Penelitian ini tidak menitikberatkan pada pengembangan algoritma baru, tetapi pada implementasi *AES* sebagai inovasi asesmen pendidikan yang mendukung guru dalam meningkatkan objektivitas, efisiensi, dan kualitas umpan balik penilaian. Melalui pendekatan ini, diharapkan hasil penelitian dapat memberikan kontribusi nyata terhadap penguatan praktik asesmen berbasis teknologi dalam pembelajaran bahasa Indonesia serta mendorong integrasi antara hasil riset *NLP* dan inovasi pendidikan di sekolah menengah.

2. METODE

Penelitian ini menggunakan desain studi kasus evaluatif dengan pendekatan kuantitatif dan kualitatif deskriptif. Pendekatan ini dipilih untuk mengkaji efektivitas penerapan *Automated Essay Scoring (AES)* sebagai inovasi asesmen menulis dalam konteks pendidikan menengah. Studi kasus dipusatkan di SMA Negeri 1 Genteng, yang menjadi lokasi uji coba penerapan sistem *AES*. Pendekatan kuantitatif digunakan untuk menganalisis kesesuaian antara skor model dan penilaian guru menggunakan metrik statistik, sedangkan pendekatan kualitatif digunakan untuk memahami persepsi guru dan siswa terhadap keadilan, kemudahan, serta manfaat sistem dalam proses pembelajaran (Fatqurhohman et al., 2025; Sidik et al., 2025). Dengan desain ini, penelitian diharapkan tidak hanya mengukur performa teknis model, tetapi juga menilai keberterimaan pedagogis inovasi teknologi dalam praktik pembelajaran menulis.

2.1 Data

Data utama penelitian ini terdiri atas kumpulan esai siswa SMA Negeri 1 Genteng yang dikumpulkan melalui tugas menulis argumentatif dan deskriptif pada mata pelajaran bahasa Indonesia. Setiap esai memiliki panjang rata-rata 200–300 kata dan diberi skor manual oleh dua guru bahasa Indonesia. Penilaian dilakukan untuk memastikan reliabilitas antarpemilai (*inter-rater reliability*) serta membentuk *ground truth* yang digunakan dalam pelatihan model. Untuk keperluan training dan validasi model, penelitian ini juga menggunakan dataset publik dari penelitian (Rajagede, 2021) berisi ribuan jawaban singkat (*short answer*) siswa berbahasa Indonesia dengan label biner (relevan/tidak relevan terhadap stimulus).

Proses pelatihan model menggunakan gabungan dataset publik dan data sekolah untuk meningkatkan kemampuan generalisasi model. Proporsi pembagian data ditetapkan sebesar 70% untuk training set, 15% untuk validation set, dan 15% untuk testing set. Data pada tahap pelatihan digunakan untuk mengajarkan pola linguistik dan struktur esai kepada model, sementara validation set digunakan untuk memantau kinerja model selama proses pelatihan agar tidak terjadi *overfitting* (Sidik et al., 2023). Adapun testing set berfungsi untuk mengukur performa akhir model terhadap data yang belum pernah dilihat sebelumnya (Sidik et al., 2023). Seluruh teks dalam dataset diproses melalui tahap *pre-processing*, meliputi tokenisasi, normalisasi huruf, penghapusan tanda baca dan karakter non-alfanumerik, serta konversi ke vektor numerik menggunakan *word embedding* pra-latih Bahasa Indonesia.

2.2 Model

Model utama yang digunakan adalah *Bidirectional Long Short-Term Memory (Bi-LSTM)*, karena arsitektur ini efektif dalam memahami konteks dua arah (*forward* dan *backward*) dalam teks sehingga dapat menangkap hubungan semantik dan sintaksis antar kalimat. Model dilatih menggunakan data yang telah diproses dengan pembagian 70:15:15 sebagaimana dijelaskan sebelumnya. Pelatihan dilakukan dengan optimasi *hyperparameter* seperti *learning rate*, jumlah unit tersembunyi, dan tingkat *dropout* untuk menghindari *overfitting*.

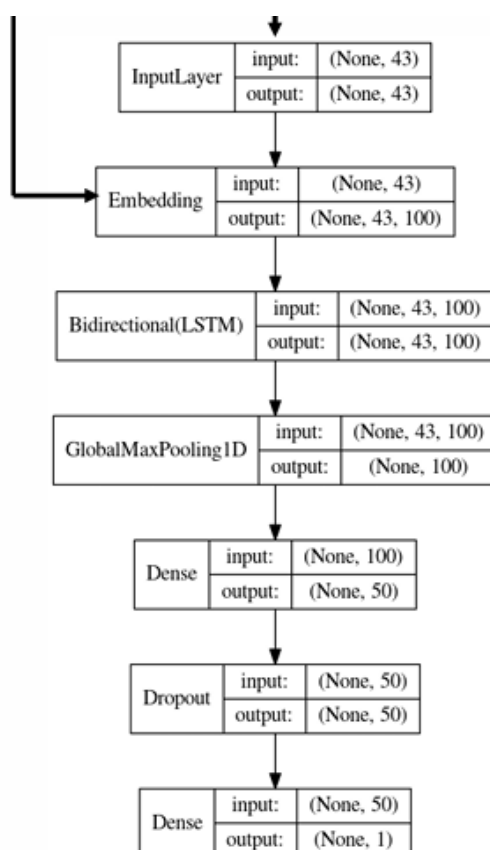


Figure 1. Model Peneletian

2.3 Evaluasi dan Analisis

Evaluasi dilakukan dengan membandingkan skor hasil prediksi model dan skor manual dari guru. Tiga metrik digunakan, yaitu *Quadratic Weighted Kappa (QWK)*, *Mean Absolute Error (MAE)*, dan *Pearson Correlation Coefficient (PCC)*. *QWK* digunakan sebagai metrik utama karena mengukur kesepakatan tertimbang antarpemilai dengan mempertimbangkan bobot kesalahan (Doewes et al., 2023). *MAE* digunakan untuk menilai deviasi rata-rata antara skor model dan skor guru, sementara *PCC* digunakan untuk melihat korelasi linear antarpemilaian (Ramesh & Sanampudi, 2022).

Table 1. Metrik Evaluasi Kinerja Model AES

Metrik Evaluasi	Deskripsi	Nilai	Interpretasi
<i>Quadratic Weighted Kappa (QWK)</i>	Mengukur tingkat kesepakatan antara skor model dan skor guru dengan mempertimbangkan bobot perbedaan antar kategori skor	$0 \leq QWK$; $QWK \leq 1$	Nilai mendekati 1 bahwa kesesuaian tinggi antara model dan penilai manusia
<i>Mean Absolute Error (MAE)</i>	Menghitung rata-rata selisih absolut antara skor prediksi model dan skor aktual dari guru.	$MAE \geq 0$	Nilai lebih kecil bahwa kesalahan prediksi yang lebih rendah dan akurasi model yang lebih tinggi
<i>Pearson Correlation Coefficient (PCC)</i>	Mengukur kekuatan dan arah hubungan linear antara skor model dan skor manusia.	$-1 \leq PCC \leq 1$	Nilai mendekati +1, ada hubungan linear positif yang kuat; Nilai 0 berarti tidak ada hubungan; Nilai -1 berarti korelasi negatif sempurna.

Analisis dilakukan dalam dua bentuk, yaitu analisis kuantitatif dan analisis kualitatif. Analisis kuantitatif dilakukan dengan menghitung nilai *QWK*, *MAE*, dan *PCC* untuk mengukur kesesuaian skor *AES* dengan penilaian guru, serta memvisualisasikan perbedaan skor model dan skor manual menggunakan diagram sebar. Sementara itu, analisis kualitatif dilakukan dengan menganalisis data persepsi guru dan siswa melalui kuesioner dan wawancara terstruktur mengenai aspek keadilan, kecepatan, dan kemudahan penggunaan sistem *AES*. Data hasil wawancara dianalisis menggunakan teknik *thematic analysis* untuk mengidentifikasi pola penerimaan terhadap inovasi asesmen berbasis AI di lingkungan sekolah (Rajagede, 2021).

3. HASIL DAN PEMBAHASAN

3.1. Hasil

Hasil penelitian yang diperoleh dari penerapan dan pengujian sistem *Automated Essay Scoring (AES)* berbasis *Bidirectional Long Short-Term Memory (Bi-LSTM)* dalam konteks penilaian tulisan siswa di SMA Negeri 1 Genteng. Pembahasan meliputi hasil evaluasi kinerja model *AES* secara kuantitatif dan analisis persepsi guru serta siswa terhadap penggunaan sistem ini dalam proses pembelajaran menulis.

Hasil Evaluasi Model AES

Model *Automated Essay Scoring (AES)* dikembangkan menggunakan arsitektur *Bidirectional Long Short-Term Memory (Bi-LSTM)* dengan input berbasis *word embedding* Bahasa Indonesia. Data pelatihan terdiri atas gabungan dataset publik dan dataset esai siswa SMA Negeri 1 Genteng, dengan pembagian data 70% untuk pelatihan, 15% validasi, dan

15% pengujian. Model dilatih selama 20 epoch dengan learning rate 0.001 dan batch size 32.

Hasil evaluasi menunjukkan bahwa model mampu memberikan skor yang sangat mendekati penilaian manual guru. Nilai *Quadratic Weighted Kappa (QWK)* sebesar 0,83, *Mean Absolute Error (MAE)* sebesar 0,41, dan *Pearson Correlation Coefficient (PCC)* sebesar 0,86 menunjukkan performa tinggi dalam meniru penilaian manusia. Nilai $QWK > 0,8$ menandakan kesesuaian yang sangat baik antara skor model dan skor guru, sedangkan nilai PCC mendekati +1 mengindikasikan korelasi yang kuat antara keduanya.

Table 2. Hasil Evaluasi Kinerja Model AES

No	Metrik Evaluasi	Nilai	Interpretasi	Kategori
1	<i>Quadratic Weighted Kappa (QWK)</i>	0,83	Kesepakatan tinggi antara skor model dan guru	Sangat baik
2	<i>Mean Absolute Error (MAE)</i>	0,41	Rata-rata selisih skor rendah	Stabil
3	<i>Pearson Correlation Coefficient (PCC)</i>	0,86	Korelasi kuat antara model dan guru	Sangat kuat

Nilai *MAE* yang rendah memperlihatkan bahwa kesalahan prediksi rata-rata hanya sekitar 0,4 poin dari skor manual, sehingga model dapat dianggap cukup stabil dan presisi. Hal ini menunjukkan bahwa *Bi-LSTM* mampu menangkap pola sintaktik dan semantik pada tulisan siswa dengan baik, termasuk dalam memahami variasi gaya bahasa, panjang kalimat, dan struktur paragraf.

Persepsi Guru dan Siswa terhadap Sistem AES

Selain hasil kuantitatif, penelitian ini juga menganalisis tanggapan guru dan siswa terhadap penerapan sistem AES melalui kuesioner dan wawancara semi-terstruktur. Sebanyak 5 guru bahasa Indonesia dan 40 siswa berpartisipasi dalam survei ini. Secara umum, mayoritas responden menunjukkan sikap positif terhadap penerapan AES. Mereka menilai sistem ini mampu meningkatkan efisiensi penilaian, memberikan umpan balik cepat, dan membantu pembelajaran menulis yang lebih reflektif. Hasil survei dirangkum pada Tabel 3 berikut.

Table 3. Persepsi Guru dan Siswa terhadap Sistem AES

No	Aspek yang Dinilai	Guru Setuju (%)	Siswa Setuju (%)	Keterangan
1	Penilaian lebih cepat dan efisien	100	90	Sistem mempercepat proses koreksi esai
2	Hasil penilaian objektif dan konsisten	80	78	Nilai tidak dipengaruhi bias antar penilai
3	Umpan balik membantu perbaikan tulisan	90	85	Feedback instan membantu refleksi belajar
4	Penilaian terhadap kreativitas dan konteks	60	55	Masih perlu intervensi guru
5	Kelayakan digunakan di sekolah	95	87	Layak sebagai alat bantu asesmen

Dari tabel tersebut terlihat bahwa 100% guru dan 90% siswa menyatakan bahwa AES mempercepat proses penilaian. Guru menyebut bahwa waktu koreksi esai yang biasanya memakan waktu 2–3 hari dapat dipangkas hingga 60% karena sistem mampu melakukan pre-scoring otomatis sebelum verifikasi manual. Sementara itu, siswa mengaku mendapatkan umpan balik lebih cepat terkait tata bahasa dan struktur tulisan, yang membantu mereka memperbaiki kesalahan secara mandiri sebelum revisi akhir. Namun, sebanyak 40% guru dan 45% siswa masih meragukan kemampuan sistem dalam menilai aspek kreatif, emosional, dan argumentatif yang sering kali bersifat subjektif. Guru juga menekankan bahwa sistem ini tidak dapat sepenuhnya memahami konteks budaya dan nuansa makna dalam tulisan siswa. Oleh karena itu, mereka menilai AES sebaiknya berfungsi sebagai alat bantu asesmen (*assessment support tool*), bukan sebagai pengganti penuh peran guru dalam menilai esai.

3.2. Pembahasan

Hasil penelitian sistem *Automatic Essay Scoring (AES)* berbasis *Bidirectional Long Short-Term Memory (Bi-LSTM)* memiliki potensi yang signifikan dalam meningkatkan efisiensi serta objektivitas proses penilaian kemampuan menulis (*writing assessment*) siswa. Model ini menunjukkan performa unggul berdasarkan hasil pengujian menggunakan metrik *Quadratic Weighted Kappa (QWK)* dan *Pearson Correlation Coefficient (PCC)*. Nilai QWK yang tinggi mencerminkan tingkat kesesuaian yang kuat antara skor yang dihasilkan oleh sistem dengan penilaian yang dilakukan oleh manusia, sementara nilai PCC yang tinggi menunjukkan adanya korelasi positif yang konsisten antara kedua jenis penilaian tersebut. Temuan ini menegaskan bahwa sistem AES berbasis Bi-LSTM mampu memberikan hasil penilaian yang akurat, stabil, dan dapat diandalkan sebagai alternatif atau pendamping penilaian manual dalam konteks evaluasi kemampuan menulis siswa.

Penerapan sistem *Automatic Essay Scoring (AES)* memberikan kontribusi nyata terhadap peningkatan efektivitas penilaian dan kualitas pengalaman belajar siswa. Penggunaan sistem ini memungkinkan guru menghemat waktu dalam menilai aspek-aspek teknis, seperti tata bahasa dan struktur kalimat, sehingga waktu pembelajaran dapat lebih difokuskan pada pengembangan kemampuan berpikir kritis, kreativitas, dan ekspresi ide siswa. Penelitian yang dilakukan oleh (Vo et al., 2023) menunjukkan bahwa integrasi AES dalam asesmen formatif mampu membantu siswa meningkatkan kualitas tulisan melalui pemberian umpan balik instan. Umpan balik yang cepat dan spesifik tersebut mendorong keterlibatan aktif, refleksi diri, serta pembelajaran yang lebih mandiri dalam proses menulis.

Sementara itu, guru dapat lebih memusatkan perhatian pada aspek-aspek penilaian yang memerlukan pemahaman mendalam dan kontekstual, seperti substansi tulisan, pengembangan argumen, kreativitas, gaya berpikir kritis, serta orisinalitas ide, sementara sistem AES menangani bagian penilaian yang bersifat mekanis dan normatif. Namun demikian, baik guru maupun siswa berpendapat bahwa tulisan yang mencakup empati, nuansa emosional, konteks sosial dalam tulisan, serta sensitivitas terhadap latar belakang budaya dan linguistik masih memerlukan keterlibatan manusia. Dengan demikian, kombinasi antara kecerdasan buatan dan penilaian manusia menjadi pendekatan paling ideal untuk memastikan proses asesmen yang tidak hanya efisien dan objektif, tetapi juga adil, kontekstual, dan berorientasi pada nilai-nilai kemanusiaan dalam pembelajaran menulis.

Penelitian (Xiao et al., 2025) menyoroti bahwa integrasi antara sistem AI dan penilaian manusia menghasilkan reliabilitas tertinggi dibandingkan penggunaan tunggal. Hal ini mendukung pandangan bahwa AES sebaiknya tidak menggantikan peran guru sepenuhnya, tetapi berfungsi sebagai asisten evaluatif cerdas (*intelligent co-assessor*). Penelitian (Semba, 2024)) di konteks pendidikan Indonesia juga menemukan bahwa sistem penilaian otomatis berbasis AI meningkatkan efisiensi kerja guru dan transparansi penilaian, meskipun aspek afektif dan sosial tetap memerlukan penilaian manusia agar hasil evaluasi lebih adil dan kontekstual.

Dengan demikian, AES bukanlah pengganti guru, tetapi pendamping cerdas dalam proses asesmen. Kombinasi antara kecerdasan buatan dan penilaian manusia akan menghasilkan sistem penilaian yang lebih adaptif, adil, dan berorientasi pada pembelajaran berkelanjutan (*lifelong learning*), sesuai dengan arah inovasi pendidikan abad ke-21.

4. KESIMPULAN

Penelitian ini bertujuan untuk mengembangkan dan mengevaluasi sistem *Automated Essay Scoring (AES)* berbasis *Bidirectional Long Short-Term Memory (Bi-LSTM)* sebagai inovasi dalam penilaian tulisan siswa di SMA Negeri 1 Genteng. Hasil penelitian menunjukkan bahwa model yang dikembangkan mampu meniru penilaian manusia dengan tingkat akurasi tinggi dan stabilitas yang baik. Nilai *Quadratic Weighted Kappa (QWK)* sebesar 0,83, *Mean Absolute Error (MAE)* sebesar 0,41, dan *Pearson Correlation Coefficient (PCC)* sebesar 0,86 menunjukkan kesesuaian kuat antara hasil penilaian model dan skor manual guru. Dengan demikian, sistem ini dinilai layak digunakan sebagai alat bantu asesmen otomatis (*assessment support tool*) dalam pembelajaran menulis Bahasa Indonesia.

Dari perspektif pendidikan, hasil kuesioner dan wawancara menunjukkan bahwa guru dan siswa memiliki persepsi positif terhadap penggunaan sistem AES. Guru menilai sistem ini efektif mempercepat proses penilaian dan memberikan umpan balik awal yang membantu mereka fokus pada pembinaan aspek isi dan kreativitas. Siswa merasa terbantu karena dapat segera mengetahui kesalahan tata bahasa dan struktur tulisan, sehingga mendorong *self-reflection* dan pembelajaran menulis yang lebih mandiri. Namun, baik guru maupun siswa menekankan pentingnya peran manusia untuk tetap menilai aspek kreatif, emosional, dan kontekstual yang belum sepenuhnya dapat diukur oleh mesin.

Secara keseluruhan, penerapan AES dalam pembelajaran menulis di sekolah menengah memiliki potensi besar untuk meningkatkan efisiensi, objektivitas, dan kualitas penilaian. Meski demikian, sistem ini tidak dimaksudkan untuk menggantikan peran guru, melainkan sebagai pendamping cerdas (*intelligent assistant*) yang membantu proses asesmen berbasis data dan umpan balik cepat. Kolaborasi antara kecerdasan buatan dan kompetensi pedagogis guru merupakan langkah strategis menuju transformasi asesmen pendidikan yang adaptif, adil, dan berorientasi pada pembelajaran berkelanjutan (*lifelong learning*) di era digital.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada SMA Negeri 1 Genteng atas dukungan, kerja sama, dan izin yang diberikan selama pelaksanaan penelitian ini. Ucapan terima kasih juga disampaikan kepada para guru bahasa Indonesia yang telah berpartisipasi sebagai penilai manual serta kepada seluruh siswa yang berkontribusi dalam penyusunan dan

pengumpulan data esai. Dukungan tersebut sangat membantu dalam menyempurnakan hasil penelitian ini.

DAFTAR PUSTAKA

- Aini, D. L., Sukmawati, K., & Amroni, T. W. (2025). Innovative Reading Aloud Approach for Enhancing Pronunciation and Speaking Fluency of Students. *JINEA: Journal of Innovation in Education and Learning*, 1(1), 9–18. <https://jurnal.ihsancahayapustaka.id/index.php/jinea>
- Chamidah, N., Yulianti, E., & Budi, I. (2023). Evaluating the Impact of Sentence Tokenization on Indonesian Automated Essay Scoring Using Pretrained Sentence Embeddings. *Revue d'Intelligence Artificielle*, 37(5), 1101–1108. <https://doi.org/10.18280/ria.370502>
- Dhini, B. F., Girsang, A. S., Sufandi, U. U., & Kurniawati, H. (2023). Automatic essay scoring for discussion forum in online learning based on semantic and keyword similarities. *Asian Association of Open Universities Journal*, 18(3), 262–278. <https://doi.org/10.1108/AAOUJ-02-2023-0027>
- Doewes, A., Kurdhi, N. A., & Saxena, A. (2023). Evaluating Quadratic Weighted Kappa as the Standard Performance Metric for Automated Essay Scoring. *International Educational Data Mining Society*, 103–113. <https://doi.org/10.5281/zenodo.8115784>
- Fatqurhohman, F., Syam, H., Puspasari, R., Niam, F., & Surur, A. M. (2025). STEM Digital Collaboration to Enhance Critical Thinking Skills of Secondary School Students: A Literature Review. *JINEA: Journal of Innovation in Education and Learning*, 1(1), 37–50. <https://jurnal.ihsancahayapustaka.id/index.php/jinea>
- Hiver, P., Al-Hoorie, A. H., Vitta, J. P., & Wu, J. (2024). Engagement in language learning: A systematic review of 20 years of research methods and definitions. *Language Teaching Research*, 28(1), 201–230. <https://doi.org/10.1177/13621688211001289>
- Hussein, M. A., Hassan, H., & Nassef, M. (2019). Automated language essay scoring systems: A literature review. *PeerJ Computer Science*, 5, e208. <https://doi.org/10.7717/peerj-cs.208>
- Jayadianti, H., Santosa, B., Cahyaning, J., Saifullah, S., & Drezewski, R. (2023). Essay auto-scoring using N-Gram and Jaro Winkler based Indonesian Typos. *MATRIK: Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, 22(2), 325–338. <https://doi.org/10.30812/matrik.v22i2.2473>
- Margareta, L., Avatus Marheni, R., & Purwitasari, H. (2025). Evaluation of Digital Learning Integration to Improve Junior High School Students' Digital Literacy. *JINEA: Journal of Innovation in Education and Learning*, 1(1), 19–28. <https://jurnal.ihsancahayapustaka.id/index.php/jinea>
- Pradani, K. A., & Suadaa, L. H. (2023). Automated Essay Scoring Menggunakan Semantic Textual Similarity Berbasis Transformer Untuk Penilaian Ujian Esai. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 10(6), 1177–1184. <https://doi.org/10.25126/jtiik.1067338>
- Rajagede, R. A. (2021). Improving Automatic Essay Scoring for Indonesian Language using Simpler Model and Richer Feature. *Kinetik: Game Technology, Information System*,

- Computer Network, Computing, Electronics, and Control*, 6(1), 11–18. <https://doi.org/10.22219/kinetik.v6i1.1196>
- Ramesh, D., & Sanampudi, S. K. (2022). An Automated Essay Scoring Systems: A Systematic Literature Review. *Artificial Intelligence Review*, 55(3), 2495–2527. <https://doi.org/10.1007/s10462-021-10068-2>
- Semba, D. L. (2024). Implementasi Long-Short Term Memory (LSTM) Untuk Generasi Feedback Berbahasa Indonesia Pada Sistem Penilaian Esai. *Jurnal FASILKOM*, 14(1), 101–113. <https://doi.org/10.37859/jf.v14i1.6906>
- Sidik, D. P., Irawijayanti, F., & Baihaqi, A. (2025). Digital Learning 5.0: Leveraging Adaptive, Immersive, and Inclusive Technologies to Overcome Educational Inequity. *JINEA: Journal of Innovation in Education and Learning*, 1(2), 75–92. <https://jurnal.ihsancahayapustaka.id/index.php/jinea>
- Sidik, D. P., Utaminingrum, F., & Muflikhah, L. (2023). Penggunaan Variasi Model pada Arsitektur EfficientNetV2 untuk Prediksi Sel Kanker Serviks. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 7(5), 2116–2121. <http://j-ptiik.ub.ac.id>
- Susilawati, E., Lubis, H., Kesuma, S., & Pratama, I. (2022). Antecedents of Student Character in Higher Education: The role of the Automated Short Essay Scoring (ASES) digital technology-based assessment model. *Eurasian Journal of Educational Research*, 2022(98). <https://doi.org/10.14689/ejer.2022.98.013>
- Verdikha, N. A., Dwiagam, J. H., & Hasudungan, R. (2024). Indonesian Automated Essay Scoring with Bag of Word and Support Vector Regression. *JSE: Journal of Science and Engineering*, 1(2), 95–100. <https://doi.org/10.30650/jse.v1i2.3841>
- Vo, Y., Rickels, H., Welch, C., & Dunbar, S. (2023). Human scoring versus automated scoring for english learners in a statewide evidence-based writing assessment. *Assessing Writing*, 56, 100719. <https://doi.org/10.1016/j.asw.2023.100719>
- Wiratmo, A., & Fatichah, C. (2020). Indonesian Short Essay Scoring Using Transfer Learning Dependency Tree LSTM. *International Journal of Intelligent Engineering and Systems*, 13(2). <https://doi.org/10.22266/ijies2020.0430.27>
- Xiao, C., Ma, W., Song, Q., Xu, S. X., Zhang, K., Wang, Y., & Fu, Q. (2025). Human-AI Collaborative Essay Scoring: A Dual-Process Framework with LLMs. *LAK '25: Proceedings of the 15th International Learning Analytics and Knowledge Conference*, 293–305. <https://doi.org/10.1145/3706468.3706507>