

PENERAPAN BERT INDONLU UNTUK AUTOMATIC ESSAY SCORING PADA MATA PELAJARAN PENDIDIKAN KEWARGANEGARAAN

Laila Nur Komaria, Muhammad Misdram, Anang Aris Widodo

Teknik Informatika, Universitas Merdeka Pasuruan

Jl. Ir. H. Juanda No.68, Tapaan, Kec. Bugul Kidul, Kota Pasuruan, Jawa Timur 67129

nurkomaria@student.unmerpas.ac.id

ABSTRAK

Penilaian jawaban esai secara manual memerlukan waktu yang relatif lama serta berpotensi menimbulkan inkonsistensi karena dipengaruhi oleh subjektivitas penilai. Permasalahan tersebut mendorong pengembangan sistem *Automatic Essay Scoring* (AES) yang mampu membantu proses penilaian secara otomatis. Penelitian ini bertujuan menerapkan model BERT IndoNLU (IndoBERT) pada sistem penilaian esai otomatis untuk mata pelajaran Pendidikan Kewarganegaraan (PKn). Metode penelitian meliputi tahap *preprocessing* teks berupa *cleaning*, *case folding*, dan *tokenization*, dilanjutkan dengan proses *fine-tuning* model IndoBERT serta evaluasi menggunakan metrik *Mean Absolute Error* (MAE). Hasil pengujian menunjukkan bahwa model menghasilkan nilai MAE sebesar 13,21, yang menunjukkan masih terdapat selisih antara skor prediksi sistem dengan skor yang diberikan oleh guru. Meskipun demikian, model telah mampu mempelajari hubungan antara jawaban siswa dan kunci jawaban sehingga dapat menghasilkan prediksi skor secara otomatis. Hasil penelitian menunjukkan bahwa model berbasis *Transformer*, khususnya IndoBERT, memiliki potensi untuk diterapkan sebagai alternatif dalam sistem penilaian esai berbahasa Indonesia. Pengembangan lebih lanjut melalui penambahan variasi data pelatihan dan optimasi parameter model diharapkan dapat meningkatkan akurasi prediksi pada penelitian berikutnya.

Kata kunci : *Automatic Essay Scoring, IndoBERT, BERT IndoNLU, Natural Language Processing, Mean Absolute Error*

1. PENDAHULUAN

Salah satu faktor penting dalam meningkatkan kualitas sumber daya manusia adalah pendidikan, sehingga untuk mengukur pencapaian kompetensi peserta didik proses evaluasi hasil belajar memiliki peran yang penting. Ujian berbentuk esai banyak digunakan karena mampu mengukur kemampuan berpikir tingkat tinggi, seperti analisis, argumentasi, dan penyampaian gagasan secara mandiri, yang tidak dapat diakomodasi secara optimal oleh soal pilihan ganda [1].

Namun, karakteristik jawaban esai yang beragam menyebabkan proses penilaian menjadi lebih kompleks dibandingkan bentuk evaluasi lainnya.

Proses penilaian esai secara manual masih menjadi tantangan bagi tenaga pendidik karena membutuhkan waktu yang cukup lama dan rentan terhadap subjektivitas. Seorang guru rata-rata memerlukan sekitar 7–10 menit untuk mengoreksi satu berkas jawaban esai dari seorang siswa [2].

Selain menurunkan efisiensi, proses koreksi manual juga berpotensi menghasilkan inkonsistensi penilaian akibat kelelahan maupun bias psikologis, sehingga objektivitas penilaian dapat berkurang [3]. Kondisi tersebut mendorong perlunya sistem penilaian otomatis (*Automatic Essay Scoring/AES*) yang mampu menghasilkan penilaian secara lebih cepat dan konsisten.

Berbagai penelitian telah mengembangkan metode *Automatic Essay Scoring* menggunakan pendekatan *Natural Language Processing* (NLP).

Akan tetapi, metode berbasis pencocokan kata kunci maupun *Cosine Similarity* masih memiliki keterbatasan dalam memahami makna semantik ketika jawaban siswa menggunakan sinonim atau penyusunan kalimat yang tidak sama dari kunci jawaban [1].

Pendekatan lain seperti *Naive Bayes* juga masih bergantung pada frekuensi kemunculan kata sehingga kurang mampu menangkap konteks kalimat secara menyeluruh [2].

Sementara itu, metode berbasis *FastText* dan *Backpropagation* telah meningkatkan representasi kata, tetapi belum optimal dalam memahami hubungan kontekstual dua arah pada kalimat yang lebih kompleks [4].

Perkembangan model berbasis *Transformer* membuka potensi baru untuk mengembangkan sistem penilaian esai otomatis. IndoBERT adalah salah satu model yang dirancang khusus untuk bahasa Indonesia, yang telah membuktikan kinerja yang sangat baik pada beragam tugas klasifikasi teks dengan tingkat akurasi mencapai 91% pada analisis sentimen [5].

Meskipun demikian, penerapan model IndoBERT pada sistem *Automatic Essay Scoring*, khususnya untuk penilaian jawaban esai siswa tingkat SMP, masih relatif terbatas sehingga menjadi peluang penelitian yang perlu dikembangkan.

Berdasarkan kondisi tersebut, penelitian ini mengusulkan penerapan model BERT IndoNLU (IndoBERT) yang telah melalui proses *fine-tuning* untuk sistem *Automatic Essay Scoring* pada mata

pelajaran Pendidikan Kewarganegaraan (PKn). Penelitian difokuskan pada jawaban esai siswa SMP dengan penilaian berdasarkan tingkat kesesuaian jawaban terhadap kunci jawaban guru (*ground truth*). metrik akurasi, *precision*, *recall*, *F1-score*, dan *Mean Absolute Error* (MAE) adalah model yg digunakan untuk evaluasi , serta dibandingkan dengan pendekatan *word embedding* tradisional. Melalui pendekatan ini, penelitian bertujuan menganalisis kinerja BERT IndoNLU, mengukur tingkat kesesuaian skor sistem dengan penilai manusia (*human grader*), serta mengevaluasi efektivitas model dalam meminimalkan kesenjangan antara penilaian manual dan otomatis. Penelitian ini diharapkan mampu memberikan kontribusi dalam pengembangan teknologi *Natural Language Processing* (NLP) untuk bahasa Indonesia serta mendukung pelaksanaan evaluasi pembelajaran yang berlangsung secara lebih efisien, objektif, dan akurat.

2. TINJAUAN PUSTAKA

2.1. Penelitian Terkait

Berbagai penelitian telah mengembangkan sistem *Automatic Essay Scoring* (AES) menggunakan beragam metode. Pendekatan berbasis *FastText* dan *Backpropagation* mampu mengenali variasi kata, tetapi masih terbatas dalam memahami konteks kalimat secara menyeluruh [4].

Metode *Cosine Similarity* efektif mengukur kemiripan vektor, namun sangat bergantung pada kesamaan kata sehingga kurang mampu mengenali sinonim atau perbedaan struktur kalimat [1][6].

Sementara itu, metode *Naive Bayes* memiliki keterbatasan karena mengasumsikan setiap kata berdiri sendiri tanpa mempertimbangkan hubungan antarkonteks[2].

Perkembangan model berbasis *Transformer*, khususnya BERT dan IndoBERT, menunjukkan peningkatan performa pada berbagai tugas *Natural Language Processing* (NLP), seperti analisis sentimen dan klasifikasi teks [5], [7], [8], [9], [10], [11][12], [13].

Namun, sebagian besar penelitian sebelumnya terfokus pada analisis sentimen social media atau ulasan pengguna[14], [15], bukan pada penilaian jawaban esai. Selain itu, penelitian mengenai aspek pengalaman pengguna sistem penilaian otomatis lebih banyak membahas antarmuka dibandingkan algoritma penilaian itu sendiri [3].

Dikarenakan hal itu, penelitian ini mengisi kesenjangan tersebut dengan menerapkan model BERT IndoNLU (IndoBERT) yang telah melalui proses *fine-tuning* untuk sistem *Automatic Essay Scoring* pada mata pelajaran Pendidikan Kewarganegaraan (PKn), sehingga diharapkan mampu menghasilkan penilaian yang lebih akurat, objektif, dan mendekati penilaian guru.

2.2. Landasan Teori

Natural Language Processing (NLP) merupakan salah satu bidang dalam kecerdasan buatan yang berfokus pada kemampuan komputer untuk memahami, menganalisis, serta mengolah bahasa alami yang digunakan oleh manusia. Dalam penerapan *Automatic Essay Scoring* (AES), NLP dimanfaatkan untuk mengubah jawaban esai menjadi representasi data yang dapat dipahami dan diproses oleh sistem komputer[1].

Sebelum memasuki tahap pemodelan, data teks terlebih dahulu menjalani proses *text preprocessing* yang mencakup *cleaning*, *case folding*, dan *tokenization*. Tahapan tersebut bertujuan menghasilkan data yang lebih konsisten, terstruktur, dan siap digunakan sebagai masukan bagi model[7].

Sistem yang dikembangkan memanfaatkan konsep *Machine Learning* dan *Deep Learning* untuk mempelajari pola hubungan antara jawaban siswa dan skor yang diberikan guru. Berbeda dengan metode konvensional, *Deep Learning* mampu mengekstraksi fitur secara otomatis sehingga lebih efektif dalam memahami hubungan semantik pada teks [4], [5].

Salah satu teknologi yang banyak dipakai adalah arsitektur *Transformer* dengan mekanisme *self-attention*, yang memungkinkan model mengerti arti kata secara dua arah (*bidirectional*) sehingga merepresentasi hasil bahasa yang lebih baik.

BERT merupakan model berbasis *Transformer* yang mampu memahami maksud kalimat secara menyeluruh melalui mekanisme *bidirectional self-attention* [7], [10], [11].

Pada penelitian ini digunakan model IndoBERT yang merupakan bagian dari proyek IndoNLU, yaitu model *pre-trained* khusus bahasa Indonesia yang telah dilatih menggunakan korpus berbahasa Indonesia sehingga lebih sesuai untuk tugas pemrosesan bahasa Indonesia [8][9].

Model tersebut kemudian disesuaikan melalui proses *fine-tuning* menggunakan dataset jawaban esai yang telah diberi skor oleh guru agar mampu beradaptasi dengan karakteristik data penelitian [5].

Konsep utama dalam sistem penilaian esai otomatis adalah *semantic similarity*, yaitu mengukur kesamaan makna antara jawaban siswa dan kunci jawaban meskipun menggunakan susunan kalimat atau pilihan kata yang berbeda [8].

Pendekatan ini menjadi keunggulan IndoBERT dibandingkan metode tradisional seperti *Cosine Similarity*, yang hanya mengukur kemiripan vektor berdasarkan kemunculan kata dan kurang mampu memahami konteks semantik secara mendalam [1].

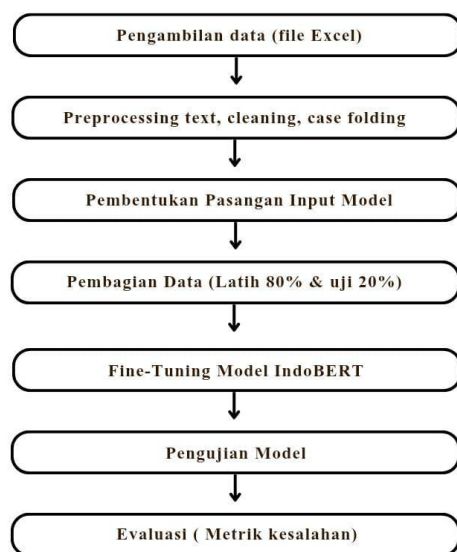
Automatic Essay Scoring (AES) merupakan sistem yang memberikan penilaian terhadap jawaban esai secara otomatis dengan tujuan meningkatkan efisiensi, konsistensi, dan objektivitas penilaian) [2], [3].

Evaluasi terhadap kinerja sistem pada penelitian ini dilakukan menggunakan beberapa metrik, yaitu *accuracy*, *precision*, *recall*, *F1-score*, dan *Mean*

Absolute Error (MAE). Di antara metrik tersebut, MAE dimanfaatkan untuk mengukur rata-rata perbedaan antara skor yang diprediksi oleh sistem dengan skor penilaian yang diberikan oleh guru. [10].

3. METODE PENELITIAN

Penelitian ini menerapkan metode eksperimen yang dilaksanakan dengan pendekatan kuantitatif. untuk mengembangkan sistem *Automatic Essay Scoring* (AES) menggunakan model BERT IndoNLU (IndoBERT). Pelaksanaan penelitian dilakukan melalui serangkaian tahapan yang mencakup proses pengumpulan data., *preprocessing* teks, pembagian dataset, proses *fine-tuning* model, dan evaluasi performa menggunakan metrik *Mean Absolute Error* (MAE). Alur penelitian secara keseluruhan ditunjukkan pada Gambar 1.



Gambar 1. Alur penelitian

3.1. Pengumpulan Data

Data yang digunakan dalam penelitian ini berupa jawaban esai mata pelajaran Pendidikan Kewarganegaraan (PKn) yang telah dinilai oleh guru sebagai *ground truth*. Skor yang diberikan guru digunakan sebagai acuan dalam proses pelatihan maupun evaluasi model. Dataset kemudian dipersiapkan dalam format yang dapat diproses oleh model BERT IndoNLU.

3.2. Preprocessing data

Tahap *preprocessing* dilakukan untuk meningkatkan kualitas data sebelum digunakan dalam proses pelatihan model. Pada penelitian ini, tahapan *preprocessing* terdiri atas beberapa proses sebagai berikut.

- Cleaning, yaitu proses menghilangkan karakter yang tidak diperlukan, seperti tanda baca tertentu, simbol, spasi berlebih, serta karakter nonalfabet yang tidak memberikan kontribusi terhadap proses pembelajaran model.

- Case Folding, yaitu proses menyeragamkan penulisan teks dengan mengonversi seluruh huruf menjadi huruf kecil (*lowercase*), sehingga variasi penulisan kata dapat diminimalkan.
- Tokenization, yaitu proses memisahkan teks menjadi token-token yang selanjutnya dikonversi menjadi representasi numerik menggunakan *tokenizer* bawaan IndoBERT agar dapat diproses oleh model.

Tahapan *preprocessing* menghasilkan data teks yang lebih terstruktur sehingga proses pelatihan model dapat berlangsung secara optimal.

3.3. Pembagian Dataset

Dataset pada penelitian ini dibagi menjadi dua bagian, yaitu *training set* dan *testing set*. *Training set* dimanfaatkan untuk proses *fine-tuning* model IndoBERT, sedangkan *testing set* digunakan untuk menguji kinerja model dalam memprediksi skor jawaban esai yang belum pernah digunakan selama proses pelatihan. Pembagian dataset dilakukan dengan rasio 80:20 agar model memperoleh data pelatihan yang cukup tanpa mengurangi kualitas proses evaluasi.

3.4. Fine-Tuning Model BERT IndoNLU

Penelitian ini memanfaatkan model BERT IndoNLU (IndoBERT), yaitu model berbasis arsitektur *Transformer* yang telah melalui proses prapelatihan (*pre-trained*) menggunakan korpus berbahasa Indonesia. penelitian ini dilakukan proses *fine-tuning* sehingga model dapat mempelajari hubungan antara jawaban siswa dengan skor yang diberikan oleh guru.

Setiap jawaban esai terlebih dahulu diproses menggunakan *tokenizer* IndoBERT untuk menghasilkan representasi token. Representasi tersebut kemudian diproses oleh lapisan encoder BERT sehingga menghasilkan *embedding* kontekstual yang mampu merepresentasikan makna kalimat secara menyeluruh. Selanjutnya, hasil representasi tersebut diteruskan ke lapisan regresi (*regression layer*) untuk menghasilkan prediksi skor jawaban esai.

Fungsi *loss* Mean Squared Error (MSE) dengan algoritma optimasi AdamW dipakai untuk pengujian. Selama proses *fine-tuning*, parameter model diperbarui secara bertahap hingga diperoleh model dengan nilai *loss* yang semakin kecil.

3.5. Evaluasi Model

Kinerja model dievaluasi menggunakan metrik *Mean Absolute Error* (MAE). Metrik ini digunakan untuk mengukur rata-rata selisih absolut antara skor hasil prediksi model dengan skor yang diberikan oleh guru. Semakin kecil nilai MAE, maka semakin baik kemampuan model dalam melakukan prediksi. Persamaan MAE ditunjukkan pada Persamaan (1).

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (1)$$

dengan:

$MAE = \text{Mean Absolute Error};$

$(n) = \text{jumlah data pengujian};$

$(y_i) = \text{skor aktual yang diberikan oleh guru};$

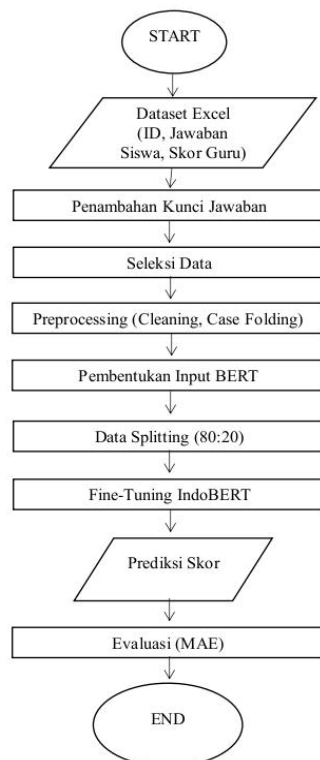
$(\hat{y}_i) = \text{skor hasil prediksi model}.$

4. HASIL DAN PEMBAHASAN

4.1. Implementasi Sistem

Sistem *Automatic Essay Scoring* (AES) dikembangkan menggunakan model BERT IndoNLU (IndoBERT) untuk melakukan penilaian otomatis terhadap jawaban esai mata pelajaran Pendidikan Kewarganegaraan (PKn). Implementasi penelitian dilakukan menggunakan bahasa pemrograman Python dengan memanfaatkan pustaka *Transformers*, *PyTorch*, dan *Scikit-learn* yang dijalankan pada lingkungan Google Colaboratory. Sistem menerima masukan berupa pasangan teks yang terdiri atas kunci jawaban dan jawaban siswa, kemudian menghasilkan prediksi skor berdasarkan hasil pembelajaran model.

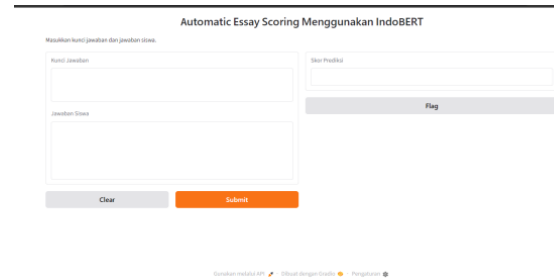
Tahapan implementasi diawali dengan proses pembacaan dataset, dilanjutkan dengan *preprocessing* berupa *cleaning*, *case folding*, dan *tokenization*. Setelah itu dataset dibagi jadi 2: data latih dan data uji menggunakan perbandingan 80:20. Data latih dipakai pada proses *fine-tuning* model IndoBERT, sedangkan data uji digunakan untuk menilai performa model menggunakan metric MAE.



Gambar 2. Flowchart implementasi sistem

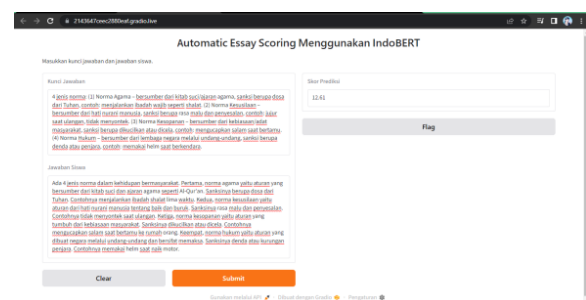
Implementasi sistem dilakukan menggunakan bahasa pemrograman Python pada Google Colab dengan memanfaatkan model IndoBERT yang telah melalui proses *fine-tuning*. Sistem menerima masukan berupa kunci jawaban dan jawaban siswa,

kemudian menghasilkan skor prediksi secara otomatis. Tampilan Antarmuka Aplikasi Automatic Essay Scoring ditunjukkan pada gambar 3



Gambar 3. Tampilan Antarmuka Aplikasi Automatic Essay Scoring

Gambar 3 menunjukkan antarmuka aplikasi Automatic Essay Scoring yang dikembangkan. Pengguna memasukkan kunci jawaban dan jawaban siswa pada kolom yang tersedia, kemudian menekan tombol Submit untuk memperoleh hasil prediksi skor berdasarkan model IndoBERT.



Gambar 4. Hasil Prediksi Skor

Pada gambar 4 menunjukkan hasil prediksi skor yang dihasilkan oleh model IndoBERT setelah sistem memproses masukan pengguna. Nilai yang ditampilkan merupakan skor prediksi yang digunakan sebagai hasil penilaian jawaban esai.

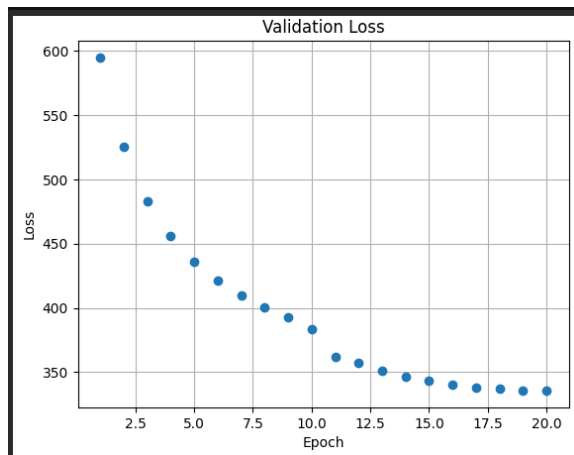
4.2. Hasil Fine-Tuning Model

Model yg digunakan untuk Proses *fine-tuning* adalah model `indobenchmark/indobert-base-p2`. Selama proses pelatihan, model mempelajari hubungan antara jawaban siswa dengan skor yang diberikan oleh guru sehingga mampu menghasilkan prediksi skor berdasarkan representasi semantik kalimat. Parameter pelatihan yang digunakan ditunjukkan pada Tabel 1.

Tabel 1. Parameter Pelatihan Model

Parameter	Nilai
Model	IndoBERT Base P2
Optimizer	AdamW
Learning Rate	2×10^{-5}
Batch Size	4
Epoch	20
Weight Decay	0,01
Loss Function	Mean Squared Error (MSE)

Selama proses pelatihan dilakukan evaluasi pada setiap *epoch* untuk memantau perubahan nilai *validation loss*. Grafik perubahan *validation loss* ditunjukkan pada Gambar 5.



Gambar 5. Grafik Validation Loss

Berdasarkan grafik yang diperoleh, nilai *validation loss* menunjukkan tren penurunan pada tahap awal pelatihan, kemudian relatif stabil hingga akhir proses pelatihan. Pola tersebut mengindikasikan bahwa model mampu mempelajari hubungan antara jawaban siswa dan skor yang diberikan oleh guru dengan baik, sehingga proses pembelajaran berlangsung secara optimal tanpa menunjukkan perubahan yang signifikan pada *epoch* berikutnya. Meskipun masih terdapat fluktuasi pada beberapa *epoch*, proses pelatihan mampu menghasilkan model yang dapat digunakan untuk melakukan prediksi terhadap data pengujian.

4.3. Hasil Evaluasi Model

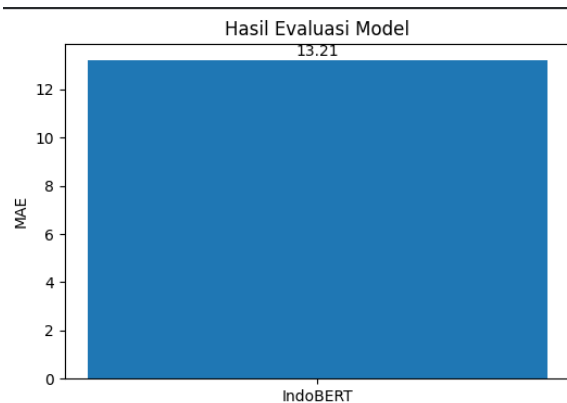
Setelah tahap training selesai, model diuji menggunakan data yang tidak digunakan selama proses *training*. Pengujian dilakukan untuk mengetahui kemampuan model dalam memberikan skor terhadap jawaban esai berdasarkan pemahaman konteks kalimat. Sebagian hasil prediksi model ditunjukkan pada Tabel 2.

Tabel 2. Hasil Prediksi Model

No	Skor Guru	Skor Prediksi	Selisih Absolut
1	23.00	14.31	8.69
2	3.00	2.70	0.30
3	22.00	14.31	7.69
4	25.00	14.32	10.68
5	6.00	5.98	0.02

Grafik MAE ditunjukkan pada Gambar 6. Berdasarkan hasil pengujian diperoleh nilai evaluasi atau Nilai MAE sebesar 13,21 menunjukkan bahwa rata-rata selisih antara skor prediksi model dan skor aktual yang diberikan oleh guru adalah sebesar 13,21 poin. Nilai tersebut menunjukkan bahwa model telah mampu melakukan prediksi terhadap jawaban esai,

namun masih terdapat perbedaan antara hasil prediksi dan penilaian guru.



Gambar 6. Grafik MAE

Perbedaan tersebut dipengaruhi oleh karakteristik jawaban esai yang memiliki variasi kosakata, struktur kalimat, serta tingkat kelengkapan informasi yang berbeda pada setiap siswa. Meskipun model IndoBERT mampu memahami hubungan semantik antarkalimat, variasi jawaban yang tinggi menyebabkan proses prediksi belum sepenuhnya mendekati skor aktual.

4.4. Pembahasan

Penerapan model BERT IndoNLU (IndoBERT) pada penelitian ini menunjukkan model mampu digunakan untuk membangun sistem *Automatic Essay Scoring* pada mata pelajaran Pendidikan Kewarganegaraan. Kemampuan utama model terletak pada representasi bahasa secara kontekstual sehingga tidak hanya mempertimbangkan kesamaan kata, tetapi juga memahami hubungan makna antar kalimat.

Pendekatan ini memberikan keunggulan dibandingkan metode berbasis pencocokan kata (*keyword matching*) karena jawaban siswa yang menggunakan sinonim atau susunan kalimat berbeda tetap dapat dipahami oleh model. Dengan demikian, proses penilaian menjadi lebih fleksibel dan mendekati cara guru dalam mengevaluasi jawaban esai.

Meskipun demikian, hasil evaluasi menunjukkan nilai MAE sebesar 13,21, yang mengindikasikan masih terdapat selisih antara hasil prediksi model dan skor aktual. Kondisi tersebut dipengaruhi oleh beberapa faktor, antara lain keterbatasan jumlah data pelatihan, variasi gaya bahasa siswa, serta kompleksitas jawaban esai yang memerlukan pemahaman konteks secara lebih mendalam. Selain itu, proses *fine-tuning* juga dipengaruhi oleh parameter pelatihan yang digunakan sehingga pemilihan parameter yang kurang optimal dapat memengaruhi performa model.

Secara keseluruhan, temuan penelitian ini menunjukkan bahwa model IndoBERT memiliki potensi yang cukup baik untuk diterapkan dalam

sistem *Automatic Essay Scoring* (AES) pada teks berbahasa Indonesia.. Pengembangan lebih lanjut melalui penambahan jumlah data pelatihan, optimasi parameter model, serta penerapan teknik augmentasi data diharapkan mampu meningkatkan performa sistem sehingga menghasilkan nilai prediksi yang lebih mendekati penilaian guru.

4.5. Kelebihan dan Keterbatasan Sistem

Berdasarkan hasil implementasi dan pengujian, sistem yang dikembangkan memiliki beberapa kelebihan, yaitu mampu melakukan penilaian jawaban esai secara otomatis, memahami hubungan semantik antar kalimat menggunakan model IndoBERT, serta meningkatkan efisiensi proses penilaian dibandingkan penilaian manual.

Di sisi lain, sistem masih memiliki beberapa keterbatasan. Nilai MAE sebesar 13,21 menunjukkan bahwa prediksi model belum sepenuhnya mendekati penilaian guru. Selain itu, performa model dipengaruhi oleh jumlah dan keberagaman dataset yang digunakan dalam proses pelatihan. Oleh karena itu, penelitian selanjutnya dapat dilakukan dengan memperbanyak dataset, melakukan optimasi *hyperparameter*, serta membandingkan performa IndoBERT dengan model berbasis *Transformer* lainnya untuk memperoleh hasil yang lebih optimal.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian, model BERT IndoNLU (IndoBERT) berhasil diimplementasikan pada sistem *Automatic Essay Scoring* untuk mata pelajaran Pendidikan Kewarganegaraan (PKn). Tahapan penelitian yang meliputi *preprocessing* data, pembentukan pasangan masukan, *fine-tuning* model, serta tahap evaluasi yang menggunakan metrik Mean Absolute Error (MAE) mengindikasikan bahwa model mampu memberikan penilaian otomatis terhadap jawaban esai dengan mempertimbangkan konteks serta makna yang terkandung dalam jawaban siswa. Meskipun masih terdapat perbedaan antara skor prediksi sistem dan skor yang diberikan oleh guru, temuan penelitian mengindikasikan bahwa pendekatan berbasis *Transformer* memiliki potensi sebagai alternatif untuk membantu proses penilaian esai menjadi lebih cepat, konsisten, dan objektif. Pada penelitian berikutnya, disarankan untuk menggunakan dataset dengan jumlah yang lebih banyak serta variasi yang lebih beragam, melakukan optimasi *hyperparameter* seperti *learning rate*, *batch size*, dan jumlah *epoch*, serta membandingkan performa IndoBERT dengan model bahasa lainnya guna meningkatkan akurasi serta kemampuan generalisasi model.

DAFTAR PUSTAKA

- [1] D. O. Sihombing, "Implementasi Natural Language Processing (NLP) dan Algoritma Cosine Similarity dalam Penilaian Ujian Esai Otomatis," *J. Sist. Komput. dan Inform.*, vol. 4, no. 2, p. 396, 2022, doi: 10.30865/json.v4i2.5374.
- [2] U. D. Mariyani, W. Setyaningsih, and R. Agustina, "Pengembangan Sistem Koreksi Jawaban Esai Otomatis Menggunakan Naive Bayes Dan Pengujian Menggunakan User Acceptance Test (UAT)," *Rainstek J. Terap. Sains dan Teknol.*, vol. 4, no. 1, pp. 61–73, 2022, doi: 10.21067/jtst.v4i1.6857.
- [3] S. Nouriska, M. C. Untoro, A. Afriansyah, M. Praseptiawan, W. Yulita, and I. F. Ashari, "User Experience Answer System Automatically With User Centered Design and User Experience Questionnaire-Short," *JITK (Jurnal Ilmu Pengetah. dan Teknol. Komputer)*, vol. 9, no. 1, pp. 81–88, 2023, doi: 10.33480/jitk.v9i1.4152.
- [4] D. A. Sani and M. Z. Sarwani, "MENGUNAKAN FASTTEXT DAN ALGORITMA BACKPROPAGATION Jurnal Nasional Pendidikan Teknik Informatika : JANAPATI | 103," vol. 11, pp. 102–111, 2022.
- [5] D. Nuryadi *et al.*, "Fine Tuning Indobert Untuk Analisis Sentimen Pada Ulasan Pengguna Aplikasi Tiket.Com Di Google Play Store," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 9, no. 2, pp. 3577–3583, 2025, doi: 10.36040/jati.v9i2.13204.
- [6] R. Tegar, A. Amin, and S. Noris, "Sistem Penilaian Esai Otomatis Menggunakan Text Embedding dan Cosine Similarity pada Platform Ujian Daring," 2026.
- [7] Widya Khafa Nofaa, Andriani Marshanda Putri, and Dewi Anggraini Puspa Hapsari, "Penerapan Metode Bert Untuk Analisis Sentimen Ulasan Pengguna Aplikasi Segari Di Google Play Store," *J. Ilm. Tek.*, vol. 4, no. 1, pp. 89–104, 2025, doi: 10.56127/juit.v4i1.1903.
- [8] P. Sayarizki and H. Nurrahmi, "Implementation of IndoBERT for Indonesian Sentiment Analysis," *Indones. J. Comput.*, vol. 9, no. August, pp. 61–72, 2024, doi: 10.34818/indoic.2024.9.2.934.
- [9] R. Merdiansah, S. Siska, and A. Ali Ridha, "Analisis Sentimen Pengguna X Indonesia Terkait Kendaraan Listrik Menggunakan IndoBERT," *J. Ilmu Komput. dan Sist. Inf.*, vol. 7, no. 1, pp. 221–228, 2024, doi: 10.55338/jikomsi.v7i1.2895.
- [10] G. Hakim, T. N. Fatyanosa, and A. W. Widodo, "Analisis Sentimen Masyarakat terhadap Kereta Cepat Whoosh pada Platform X menggunakan IndoBERT," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 8, no. 10, pp. 1–10, 2024, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [11] M. G. Al-Kadzim, R. Rasim, and H. Herbert, "Analisis Perubahan Sentimen Publik di Media Sosial X terhadap Konflik Palestina-Israel Menggunakan Model IndoBERT," *Digit. Transform. Technol.*, vol. 4, no. 2, pp. 1167–1174, 2025, doi: 10.47709/digitech.v4i2.5312.

- [12] I. G. Bagus, S. Wibawa, I. N. E. Indrayana, M. Pasek, and A. Ariawan, "PENERAPAN METODE INDOBERT UNTUK DETEKSI BERITA HOAKS PADA MEDIA DIGITAL BERBAHASA INDONESIA," vol. 07, no. 02, pp. 326–333, 2026.
- [13] G. Salwa, "EduTIK: Jurnal Pendidikan Teknologi Informasi dan Komunikasi Volume 1 Nomor 5, Oktober 2021," *EduTIK J. Pendidik. Teknol. Inf. dan Komun.*, vol. 2, no. 5, p. 773, 2025.
- [14] N. Putu, D. Agustina, I. D. Ayu, P. Pratiwi, I. G. Ngurah, and L. Wijayakusuma, "Public Sentiment Analysis on Demonstration Actions Using IndoBERT Based on Transfer Learning," vol. 9, no. 6, 2026.
- [15] I. Mursidah, R. Sanjaya, B. Yulianto, D. Sweetania, and P. Sularsih, "Klasifikasi Sentimen Google Play Store Aplikasi ChatGPT Berbahasa Indonesia Berbasis IndoBERT," *J. Minfo Polgan*, vol. 14, no. 2, pp. 3349–3359, 2025, doi: 10.33395/jmp.v14i2.15751.