E-ISSN: 2986-7258
P-ISSN : 2986-724X

PISCES Volume 04 Nomor 01 Maret 2026

Optimasi Model BiLSTM Berbasis FastText pada Data Augmentasi Semantik IndoBERT untuk Klasifikasi Teks Bahasa Indonesia

[1*]Nur Fadilah, [2]Bayu Anugerah Putra, [3]Muh. Isbar Pratama
[1,3]Universitas Negeri Makassar
[2]Universitas Muhammadiyah Riau
Email: nurfadilah@unm.ac.id[1*], bayuanugerahputra@umri.ac.id[2], isbarpratama@unm.ac.id[3]

Received : 03 Januari 2026
Accepted : 15 Februari 2026
Published : 30 Maret 2025

ABSTRAK

Evaluasi kognitif melalui tes uraian singkat memerlukan proses penilaian yang konsisten dan objektif, namun pemeriksaan manual sering menghadapi kendala berupa keterbatasan waktu dan perbedaan penilaian antar pemeriksa. *Automatic Essay Scoring* (AES) merupakan salah satu pendekatan yang dapat digunakan untuk mengotomatisasi proses penilaian tersebut. Penelitian ini mengusulkan optimasi model *Bidirectional Long Short-Term Memory* (BiLSTM) berbasis FastText pada data hasil augmentasi semantik IndoBERT untuk klasifikasi teks bahasa Indonesia. Data latih diperoleh dari hasil augmentasi EDA_Sinonim_IndoBERT pada dataset UKARA, sedangkan data validasi dan data pengujian menggunakan data asli tanpa augmentasi. Proses optimasi dilakukan melalui penerapan *Global Max Pooling* untuk meningkatkan kualitas representasi fitur serta *class weighting* untuk mengurangi bias akibat ketidakseimbangan kelas. Hasil eksperimen menunjukkan bahwa model yang diusulkan mencapai akurasi sebesar 93,490% pada data validasi dan 78,00% pada data pengujian independen. Perbedaan performa antara data validasi dan data pengujian menunjukkan bahwa meskipun augmentasi semantik mampu meningkatkan variasi data latih, kemampuan generalisasi model terhadap data yang belum pernah dilihat sebelumnya masih menjadi tantangan. Selain itu, penerapan *class weighting* mampu meningkatkan kemampuan model dalam mengenali kelas minoritas dengan nilai recall sebesar 92%, Hasil penelitian menunjukkan bahwa optimasi arsitektur dan strategi pelatihan memiliki peran penting dalam meningkatkan performa sistem *Automatic Essay Scoring* Bahasa Indonesia.

Kata Kunci: *Automatic Essay Scoring*, BiLSTM, FastText, IndoBERT, Data Augmentation

ABSTRACT

Cognitive assessment through short-answer essays requires a consistent and objective scoring process; however, manual evaluation often suffers from time constraints and inter-rater variability. Automatic Essay Scoring (AES) has emerged as a promising approach to automate the assessment process. This study proposes an optimized Bidirectional Long Short-Term Memory (BiLSTM) model combined with FastText embeddings for Indonesian text classification using semantically augmented data generated by IndoBERT. The training dataset was obtained through the EDA_Synonym_IndoBERT augmentation technique on the UKARA dataset, while the validation and testing datasets consisted of original, non-augmented responses. Model optimization was achieved through the integration of Global Max Pooling to enhance feature representation and class weighting to mitigate class imbalance. Experimental results show that the proposed model achieved an accuracy of 93.49% on the validation set and 78.00% on the independent test set. The performance gap between validation and testing results indicates that, although semantic augmentation increases the diversity of training data, model generalization to previously unseen data remains a challenging issue. Furthermore, the implementation of class weighting improved the model's ability to recognize minority-class instances, achieving a recall score of 92%. These findings demonstrate that architectural optimization and training strategies play a crucial role in improving the performance of Automatic Essay Scoring systems for the Indonesian language.

Keywords: *Automatic Essay Scoring, BiLSTM, FastText, IndoBERT, Data Augmentation*

1. PENDAHULUAN

Efektivitas evaluasi kognitif melalui tes uraian singkat sangat bergantung pada konsistensi dan objektivitas instrumen penilaian. Aliyah dkk. (2025) menegaskan bahwa tes uraian singkat merupakan instrumen krusial untuk mengukur kedalaman artikulasi argumen siswa secara presisi. Namun, implementasi manual di Indonesia masih menghadapi hambatan berupa beban kognitif pemeriksa yang tinggi serta variabilitas antar-penilai yang signifikan. Fadilah dan SG Zain (2024) mengidentifikasi bahwa keterbatasan waktu pemeriksaan menjadi determinan utama lambatnya integrasi tes uraian ke dalam sistem evaluasi berbasis komputer. Kondisi ini memicu urgensi pengembangan sistem *Automatic Essay Scoring* (AES) yang mampu mereplikasi ketepatan penilaian manusia secara otomatis.

Transformasi arsitektur AES saat ini bergeser dari metode berbasis kemiripan leksikal menuju model pembelajaran mendalam (*deep learning*) yang mengeksploitasi dependensi sekuensial teks. Implementasi *Bidirectional Long Short-Term Memory* (BiLSTM) menjadi standar karena kemampuannya menangkap konteks dua arah, yang sangat krusial dalam memahami aliran semantik pada jawaban esai pendek. Penggunaan *word embedding* FastText memperkuat model ini melalui representasi tingkat *subword*, yang secara efektif menangani masalah *Out-of-Vocabulary* (OOV) pada bahasa Indonesia dengan morfologi kompleks. Selain itu, integrasi fitur kontekstual dari model bahasa pra-latih seperti BERT terbukti meningkatkan performa secara signifikan dibandingkan model *stacking* tradisional (Fadilah dan Priyanta, 2022).

Penelitian terdahulu oleh Fadilah dan Priyanta (2022) berhasil mengidentifikasi efektivitas teknik *Easy Data Augmentation* (EDA) dalam memperkaya volume data latih pada dataset UKARA. Dalam studi tersebut, varian *EDA_Sinonim_IndoBERT* ditemukan sebagai metode augmentasi dengan performa terbaik, khususnya pada dataset dengan kompleksitas linguistik tinggi (Dataset A). Keberhasilan ini menempatkan augmentasi kontekstual sebagai solusi primer dalam mengatasi kelangkaan dataset AES bahasa Indonesia yang sering kali menghambat konvergensi model saraf. Karakteristik Dataset A sebagai *best-performing baseline* ini diposisikan sebagai tolok ukur utama (*benchmark*) dalam penelitian ini guna menguji batas atas kemampuan (*upper bound*) dari arsitektur optimasi yang diusulkan.

Namun, terlepas dari capaian tersebut, teknik augmentasi yang ada masih menyisakan masalah fundamental terkait kualitas data sintetis dan kapasitas generalisasi model. Operasi acak pada augmentasi sering kali menghasilkan *noise* semantik yang merusak integritas logika kalimat, sehingga mengaburkan esensi jawaban siswa. Sebagai contoh, penggantian kata kunci secara non-kontekstual dapat mengubah konten serta konteks asli dari informasi dalam teks ringkas (Rahmawati & Herwanto, 2026). Penggunaan kata yang tidak baku atau manipulasi acak juga rentan menghasilkan teks sintetis berisik yang memicu pengetatan *overfitting* pada model (Rahma & Suadaa, 2024). Selain itu, distribusi label yang tidak seimbang pasca-augmentasi memicu bias prediksi, di mana model cenderung gagal mengenali variasi pola pada kelas minoritas (Siregar & Sitompul, 2025). Ketidakmampuan model dalam memitigasi data sintetis yang berisik (*noisy*) dan tidak seimbang (*imbalanced*) menjadi hambatan utama dalam mencapai reliabilitas penilaian yang setara dengan penilai manusia.

Sejalan dengan tantangan tersebut, penelitian ini mereposisi fokus pada pengembangan model BiLSTM-FastText yang dirancang untuk meningkatkan *robustness* terhadap *noise* semantik hasil augmentasi IndoBERT. Kelebihan utama FastText terletak pada representasi tingkat *subword*, yang dinilai mampu menjaga ketahanan model dalam mengenali token-token anomali atau variasi kata tidak baku akibat manipulasi acak pada teks sintetis. Di sisi lain, kemampuan pemrosesan dua arah dari arsitektur BiLSTM secara efektif memitigasi distorsi makna dengan cara mengeksploitasi konteks sekuensial sebelum dan sesudah token yang berisik (*noisy*), sehingga esensi kalimat secara holistik tetap dapat dipertahankan. Implementasi pengklasifikasi sekuensial ini bertindak sebagai lapisan stabilisasi pasca-augmentasi untuk mengembalikan keandalan prediksi model.

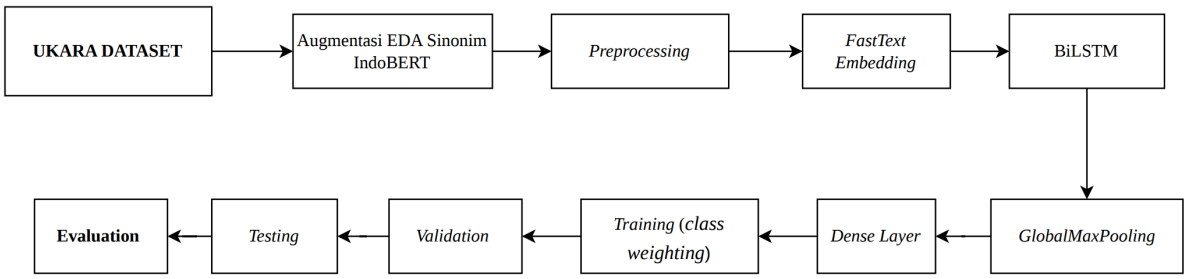
Kontribusi kebaruan (*novelty*) penelitian ini terletak pada integrasi tiga strategi optimasi: (1) penerapan mekanisme *class weighting* untuk menangani ketidakseimbangan data secara stokastik, (2) penggunaan ekstraksi fitur global untuk memperkuat representasi semantik pada teks pendek, dan (3)

skema evaluasi yang lebih ketat melalui pemisahan set validasi dan testing secara independen. Tujuan utama penelitian ini adalah mengoptimalkan ketangguhan model AES dalam menghasilkan prediksi skor yang adil dan konsisten pada dataset UKARA, sekaligus menutup celah reliabilitas yang ditinggalkan oleh metode augmentasi kontekstual sebelumnya

2. METODE PENELITIAN

2.1 Desain Penelitian

Penelitian ini mengusulkan optimasi model *Automatic Essay Scoring* (AES) berbasis *Bidirectional Long Short-Term Memory* (BiLSTM) dan FastText pada data hasil augmentasi semantik IndoBERT. Fokus utama penelitian diarahkan pada peningkatan kemampuan generalisasi model terhadap data hasil augmentasi yang berpotensi mengandung *noise* semantik dan distribusi kelas yang tidak seimbang. Tahapan penelitian terdiri atas proses persiapan dataset, *preprocessing* teks, representasi kata menggunakan FastText, pembangunan model BiLSTM, penerapan strategi optimasi model, serta evaluasi performa menggunakan data validasi dan data pengujian yang dipisahkan secara independen. Alur penelitian yang diusulkan ditunjukkan pada Gambar 1.



Gambar 1. Arsitektur Penelitian

2.2 Data Set Penelitian

Dataset yang digunakan berasal dari UKARA (Ujian Komputer Bahasa Indonesia), yang terdiri atas jawaban uraian singkat siswa beserta label penilaiannya. Data terdiri dari data A dan data B. berikut adalah distribusi datanya:

Tabel 1. Sebaran Data UKARA

Data	Jumlah Data	Label Benar	Label Salah
Training_A	854	609	245
Val_A	215	153	62
Test_A	268	191	77
Training_B	974	531	437
Val_B	244	135	109
Test_B	305	168	137

Data di atas kemudian dianalisis pada penelitian Fadilah dan Priyanta (2022). Hasil penelitian tersebut menunjukkan bahwa metode EDA_Sinonim_IndoBERT pada Dataset A menghasilkan performa tertinggi dengan akurasi sebesar 84,70% dibandingkan seluruh kombinasi metode augmentasi dan dataset yang diuji. Temuan tersebut mengindikasikan bahwa Dataset A dengan augmentasi EDA_Sinonim_IndoBERT merupakan konfigurasi terbaik (*best-performing baseline*) yang telah terverifikasi pada penelitian sebelumnya. Oleh karena itu, penelitian ini menggunakan Dataset A hasil augmentasi EDA_Sinonim_IndoBERT sebagai dasar eksperimen untuk mengevaluasi efektivitas strategi optimasi model yang diusulkan. Pendekatan ini dipilih agar peningkatan atau penurunan performa yang diperoleh dapat dikaitkan secara langsung dengan perubahan arsitektur dan strategi pelatihan, bukan dipengaruhi oleh variasi metode augmentasi maupun karakteristik dataset yang berbeda.

Selain itu, penggunaan baseline terbaik memungkinkan penelitian ini berfokus pada evaluasi kemampuan model dalam memanfaatkan data hasil augmentasi semantik secara optimal. Dengan demikian, penelitian tidak diarahkan untuk membandingkan kembali teknik augmentasi yang telah dievaluasi sebelumnya, melainkan untuk menginvestigasi sejauh mana integrasi *Global Max Pooling, class weighting*, dan skema evaluasi independen mampu meningkatkan kemampuan generalisasi model pada konfigurasi yang telah terbukti memberikan performa terbaik.

Kemudian, Data validasi dan data pengujian menggunakan data asli tanpa augmentasi untuk memastikan evaluasi model dilakukan secara objektif terhadap data yang belum pernah dilihat selama proses pelatihan. Sementara itu, data validasi dan data pengujian menggunakan data asli tanpa augmentasi untuk memastikan evaluasi kemampuan generalisasi model dilakukan secara objektif. Berikut adalah sebaran data A setelah augmentasi:

Tabel 2. Sebaran Data Setelah Augmentasi EDA Sinonim IndoBERT (Fadilah dan Priyanta, 2022)

Model	Data A		
	Total	Benar	Salah
EDA_Sinonim_Indobert	1624	1188	436

2.3 *Prepocessing* Teks

Sebelum masuk ke tahap pemodelan, seluruh data baik data latih hasil augmentasi maupun data validasi dan pengujian asli melewati tahap *preprocessing* yang sama untuk memastikan konsistensi format data. Tahap *preprocessing* ini meliputi pembersihan teks dengan menghapus tanda baca, transformasi huruf menjadi huruf kecil (*lowercase*), tokenisasi, penghapusan *stopword*, serta penyamaan panjang urutan (*padding sequence*). Hasil *preprocessing* selanjutnya digunakan sebagai masukan pada proses representasi kata (Dhini dkk., 2023)

2.4 Representasi Kata Menggunakan FastText

Representasi kata dilakukan menggunakan model FastText bahasa Indonesia pra-latih (cc.id.300.bin) dengan dimensi vektor sebesar 300. FastText dipilih karena mampu membentuk representasi berbasis subword sehingga lebih efektif dalam menangani kata yang tidak ditemukan pada kosakata pelatihan (*Out-of-Vocabulary*/OOV). Karakteristik ini sangat relevan untuk bahasa Indonesia yang memiliki keragaman morfologi, penggunaan afiks yang tinggi, serta variasi penulisan kata. Dengan memanfaatkan representasi sub-kata, FastText tetap mampu menghasilkan vektor yang representatif untuk kata baru maupun kata yang mengalami kesalahan pengetikan. Keunggulan tersebut telah dibuktikan oleh Ariyus dkk. (2024), yang menunjukkan bahwa FastText memiliki kemampuan yang baik dalam mempertahankan performa klasifikasi pada teks bahasa Indonesia yang bersifat informal dan memiliki variasi linguistik yang tinggi.

2.5 Arsitektur Model BiLSTM-FastText

Arsitektur model yang diusulkan terdiri atas lapisan embedding, *Bidirectional Long Short-Term Memory* (BiLSTM), *Global Max Pooling,* dan *Fully Connected Layer*. Lapisan embedding memanfaatkan bobot FastText yang telah dibangun sebelumnya untuk merepresentasikan setiap token ke dalam ruang vektor berdimensi 300. Selanjutnya, BiLSTM digunakan untuk mempelajari hubungan kontekstual dua arah antar kata sehingga informasi dari kata sebelumnya maupun sesudahnya dapat dimanfaatkan secara bersamaan dalam proses klasifikasi.

Keluaran BiLSTM kemudian diproses menggunakan *Global Max Pooling* untuk mengekstraksi fitur-fitur paling dominan dari representasi sekuens yang dihasilkan. Mekanisme ini membantu mempertahankan informasi penting sekaligus menghasilkan representasi fitur yang lebih ringkas dan efektif. Penggunaan strategi *pooling* global pada tugas klasifikasi teks juga telah dilaporkan mampu meningkatkan efisiensi representasi fitur tanpa mengurangi kemampuan model dalam mempertahankan

informasi semantik yang relevan (Mohammad, 2025). Hasil ekstraksi fitur selanjutnya diteruskan ke *Fully Connected Layer* dengan fungsi aktivasi ReLU untuk melakukan transformasi nonlinier terhadap fitur yang diperoleh. Pada tahap akhir, lapisan keluaran menggunakan fungsi aktivasi sigmoid untuk menghasilkan probabilitas klasifikasi biner yang digunakan dalam proses penentuan kelas jawaban.

2.6 Strategi Optimasi Model

Penelitian ini menerapkan mekanisme class weighting untuk meningkatkan stabilitas model terhadap distribusi data yang tidak seimbang. Strategi tersebut memberikan bobot yang lebih besar pada kelas minoritas sehingga mengurangi kecenderungan model untuk memprioritaskan kelas mayoritas. Selain itu, teknik *Early Stopping* digunakan untuk mencegah *overfitting* dengan menghentikan proses pelatihan secara otomatis ketika performa pada data validasi tidak lagi mengalami peningkatan setelah sejumlah *epoch* tertentu (Al-Sultan dkk., 2025).

2.7 Evaluasi Model

Evaluasi model dilakukan menggunakan data validasi dan data pengujian yang dipisahkan secara independen. Pemisahan ini bertujuan untuk memperoleh gambaran yang lebih realistis mengenai kemampuan generalisasi model. Kinerja model diukur menggunakan *Accuracy, Precision, Recall,* dan *F1-Score.* Mengingat distribusi data yang tidak seimbang, metrik *Recall* dan *F1-Sc*ore digunakan sebagai indikator utama dalam menilai performa model karena mampu memberikan gambaran yang lebih representatif terhadap kemampuan klasifikasi pada setiap kelas.

3. HASIL DAN PEMBAHASAN

3.1 Distribusi Dataset

Setelah proses augmentasi menggunakan metode EDA_Sinonim_IndoBERT, jumlah data latih meningkat dibandingkan dataset asli. Augmentasi dilakukan untuk memperkaya variasi data dan mengurangi ketidakseimbangan distribusi kelas yang umum ditemukan pada tugas *Automatic Essay Scoring* (AES). Data validasi dan data pengujian tetap menggunakan data asli tanpa augmentasi untuk memastikan evaluasi model dilakukan pada data yang merepresentasikan kondisi nyata.

Tabel 3 Distribusi Dataset yang Digunakan dalam Penelitian.

Data	Jumlah Data	Label Benar	Label Salah
Training (EDA_Sinonim_Indobert)	1624	1188	436
Validation	215	153	62
Testing	268	191	77

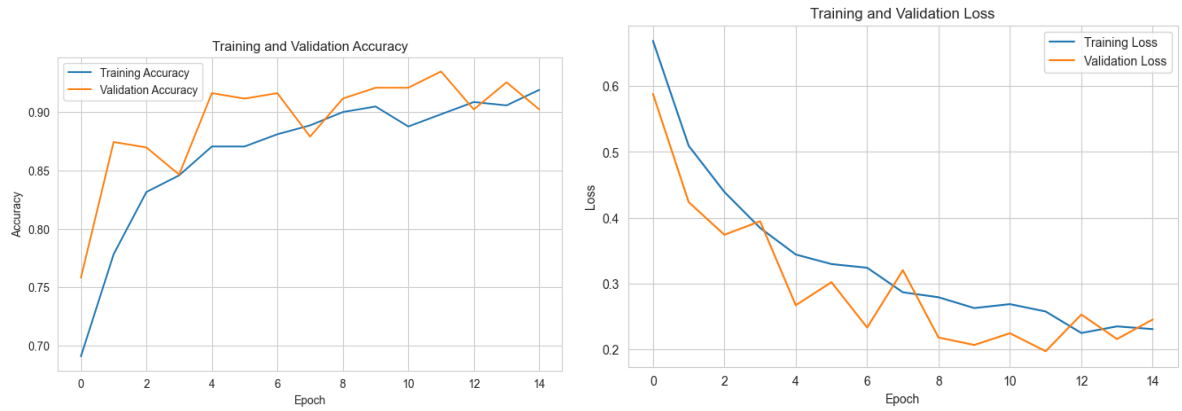
Hasil augmentasi menghasilkan distribusi kelas yang lebih seimbang dibandingkan dataset awal. Kondisi ini diharapkan mampu membantu model dalam mempelajari karakteristik masing-masing kelas secara lebih baik serta mengurangi kecenderungan model untuk memprediksi kelas mayoritas.

3.2 Hasil Pelatihan Model

Model BiLSTM-FastText dilatih menggunakan data hasil augmentasi dengan menerapkan mekanisme *class weighting* dan *Early Stopping*. Gambar 2 menunjukkan perkembangan nilai akurasi dan loss selama proses pelatihan.

E-ISSN: 2986-7258
P-ISSN : 2986-724X

PISCES Volume 04 Nomor 01 Maret 2026



Gambar 2. Perkembangan Nilai Akurasi dan *Loss* Pada Proses Pelatihan

Gambar 2 menunjukkan perkembangan nilai akurasi dan *loss* selama proses pelatihan model. Kurva akurasi menunjukkan peningkatan yang konsisten pada data pelatihan maupun data validasi. Performa terbaik diperoleh pada *epoch* ke-12 dengan nilai *validation accuracy* sebesar 93,49%. Sementara itu, kurva *loss* menunjukkan tren penurunan yang stabil hingga mencapai nilai *validation loss minimum* sebesar 0,1974. Kedekatan antara kurva pelatihan dan validasi mengindikasikan bahwa model mampu mempelajari pola jawaban siswa secara efektif tanpa mengalami *overfitting* yang signifikan

3.3 Evaluasi pada Data Validasi

Evaluasi pertama dilakukan menggunakan data validasi yang tidak terlibat secara langsung dalam proses pelatihan. Hasil klasifikasi pada data validasi ditunjukkan pada Tabel 4.

Tabel 4. Hasil Evaluasi Data Validasi

Kelas	Precision	Recall	F1-Score
0	0,86	0,92	0,89
1	0,97	0,94	0,95
Macro Avg	0,92	0,93	0,92
Weighted Avg	0,94	0,93	0,94
***Accuracy* = 93%**			

Model yang diusulkan memperoleh akurasi sebesar 93% pada data validasi dengan nilai *macro F1-score* sebesar 92% dan *weighted F1-score* sebesar 94%. Hasil tersebut menunjukkan bahwa model mampu melakukan klasifikasi secara konsisten pada kedua kelas yang diuji. Pada kelas 0, model memperoleh *precision* sebesar 86%, *recall* sebesar 92%, dan *F1-score* sebesar 89%. Sementara itu, pada kelas 1 diperoleh precision sebesar 97%, *recall* sebesar 94%, dan *F1-score sebesar 95%*. Nilai recall yang relatif tinggi pada kedua kelas menunjukkan bahwa model mampu mengenali sebagian besar sampel dengan benar tanpa menunjukkan bias yang berlebihan terhadap salah satu kelas.

Selain itu, perbedaan yang relatif kecil antara nilai macro *F1-score* dan weighted *F1-score* mengindikasikan bahwa mekanisme class weighting yang diterapkan selama proses pelatihan berhasil membantu model dalam menangani ketidakseimbangan distribusi kelas. Temuan ini menunjukkan bahwa kombinasi FastText, BiLSTM, *Global Max Pooling,* dan *class weighting* mampu menghasilkan representasi fitur yang efektif pada data hasil augmentasi semantik.

3.4 Evaluasi pada Data Pengujian

Evaluasi kemampuan generalisasi model dilakukan menggunakan data pengujian yang sepenuhnya terpisah dari proses pelatihan dan validasi. Hasil evaluasi ditunjukkan pada Tabel 5.

Tabel 5. Hasil Evaluasi Data Pengujian

Kelas	Precision	Recall	F1-Score
0	0,58	0,79	0,67
1	0,90	0,77	0,83
Macro Avg	0,74	0,78	0,75
Weighted Avg	0,81	0,78	0,79
Accuracy = 78%			

Model yang diusulkan memperoleh akurasi sebesar 78% pada data pengujian dengan nilai macro *F1-score* sebesar 75% dan *weighted F1-score* sebesar 79%. Hasil ini menunjukkan bahwa model masih mampu mempertahankan performa klasifikasi yang cukup baik ketika dihadapkan pada data yang belum pernah dilihat sebelumnya.Pada kelas 0, model memperoleh *precision* sebesar 58%, *recall* sebesar 79%, dan *F1-score* sebesar 67%. Sementara itu, pada kelas 1 diperoleh precision sebesar 90%, *recall* sebesar 77%, dan F1-score sebesar 83%. Nilai *recall* yang relatif tinggi pada kelas 0 menunjukkan bahwa model masih mampu mengenali sebagian besar sampel pada kelas tersebut, meskipun nilai *precision* yang lebih rendah mengindikasikan adanya kecenderungan model menghasilkan prediksi positif yang tidak selalu sesuai dengan label sebenarnya.

Dibandingkan dengan hasil validasi, terjadi penurunan performa pada data pengujian. Kondisi ini mengindikasikan bahwa distribusi data hasil augmentasi yang digunakan selama proses pelatihan belum sepenuhnya merepresentasikan karakteristik data asli pada tahap pengujian. Meskipun demikian, penerapan *class weighting* dan *Global Max Pooling* masih mampu membantu model mempertahankan kemampuan klasifikasi pada kedua kelas sehingga menghasilkan performa yang relatif seimbang pada data pengujian independen.

3.5 Analisis Generalisasi Model

Salah satu fokus utama penelitian ini adalah mengevaluasi kemampuan generalisasi model terhadap data yang belum pernah dilihat sebelumnya. Perbandingan performa pada data validasi dan data pengujian ditunjukkan pada Tabel 6.

Tabel 6. Perbandingan Performa Validasi dan Pengujian

Dataset	Accuracy
Validation	93%
Testing	78%

Terlihat adanya selisih performa antara data validasi dan data pengujian. Temuan ini menunjukkan bahwa peningkatan jumlah data melalui augmentasi semantik tidak secara otomatis menjamin peningkatan kemampuan generalisasi model. Hasil ini sejalan dengan temuan penelitian sebelumnya yang menunjukkan bahwa augmentasi berbasis sinonim masih berpotensi menghasilkan perubahan makna yang tidak sepenuhnya sesuai dengan konteks jawaban asli. Akibatnya, model dapat mempelajari pola yang kurang representatif terhadap distribusi data nyata.

Meskipun demikian, penerapan ekstraksi fitur global melalui *Global Max Pooling* dan mekanisme *class weighting* terbukti membantu model mempertahankan performa klasifikasi pada data pengujian. Dengan demikian, pendekatan yang diusulkan mampu meningkatkan stabilitas model terhadap variasi data hasil augmentasi dibandingkan penggunaan arsitektur BiLSTM standar.

3.6 Perbandingan dengan Penelitian Sebelumnya

Penelitian sebelumnya oleh Fadilah dan Priyanta (2022) menunjukkan bahwa metode EDA_Sinonim_IndoBERT memberikan performa terbaik dibandingkan teknik augmentasi lainnya

pada dataset UKARA. Penelitian tersebut berfokus pada pencarian metode augmentasi yang paling efektif untuk meningkatkan performa model klasifikasi. Berbeda dengan penelitian sebelumnya, penelitian ini memfokuskan perhatian pada kemampuan model dalam memanfaatkan data hasil augmentasi semantik melalui penerapan strategi optimasi berupa *class weighting*, ekstraksi fitur global, serta evaluasi menggunakan data validasi dan data pengujian yang dipisahkan secara independen.

Meskipun hasil yang diperoleh tidak dapat dibandingkan secara langsung karena perbedaan skenario evaluasi, pendekatan yang diusulkan memberikan gambaran yang lebih realistis mengenai kemampuan generalisasi model terhadap data yang belum pernah dilihat sebelumnya. Temuan ini menunjukkan bahwa optimasi arsitektur dan strategi pelatihan memiliki peran penting dalam meningkatkan kualitas sistem *Automatic Essay Scoring* berbasis bahasa Indonesia. Perbedaan karakteristik kedua penelitian dapat dilihat pada Tabel 7.

Tabel 7. Perbandingan Karakteristik Penelitian

Aspek	Fadilah & Priyanta (2022)	Penelitian Ini
Dataset	UKARA	UKARA
Fokus Penelitian	Evaluasi teknik augmentasi data	Optimasi model pada data hasil augmentasi
Metode Augmentasi	EDA, EDA Delete, EDA Swap, EDA Insert, EDA_Sinonim_IndoBERT, EDA_Sinonim_IndoBERT_Tag	EDA_Sinonim_IndoBERT
Word Embedding	FastText	FastText
Model Klasifikasi	BiLSTM	BiLSTM + GlobalMaxPooling
Penanganan Imbalance	Tidak diterapkan secara khusus	Class Weighting
Strategi Pelatihan	Pelatihan standar	Class Weighting dan Early Stopping
Data Validasi Terpisah	Tidak	Ya
Data Testing Independen	Tidak	Ya
Fokus Evaluasi	Akurasi metode augmentasi	Kemampuan generalisasi model
Kontribusi Utama	Menentukan metode augmentasi terbaik	Meningkatkan stabilitas dan generalisasi model terhadap data hasil augmentasi

Berdasarkan Tabel 7, terlihat bahwa penelitian ini merupakan pengembangan lanjutan dari penelitian sebelumnya. Jika penelitian terdahulu berfokus pada peningkatan kualitas data melalui augmentasi, maka penelitian ini berupaya meningkatkan kemampuan model dalam memanfaatkan data hasil augmentasi tersebut secara lebih efektif. Adapun perbandingan hasil eksperimen dapat dilihat pada Tabel 8.

Tabel 8. Perbandingan Hasil Penelitian

Penelitian	Model	Augmentasi	Akurasi
Fadilah & Priyanta (2022)	BiLSTM + FastText	EDA_Sinonim_IndoBERT	82,83%
Penelitian Ini (Validasi)	BiLSTM + FastText + GlobalMaxPooling + Class Weighting	EDA_Sinonim_IndoBERT	93,49%

Penelitian Ini (Testing)	BiLSTM + FastText + GlobalMaxPooling + Class Weighting	EDA_Sinonim_IndoBERT	78,00%

Secara numerik, model yang diusulkan mampu mencapai akurasi sebesar 93,49% pada data validasi. Namun demikian, performa pada data pengujian independen berada pada angka 78%. Perbedaan ini menunjukkan bahwa proses augmentasi semantik mampu membantu model mempelajari pola data selama pelatihan, tetapi masih terdapat tantangan dalam mempertahankan performa ketika model dihadapkan pada distribusi data yang berbeda. Perlu diperhatikan bahwa hasil pada Tabel 8 tidak dapat dibandingkan secara langsung karena kedua penelitian menggunakan skenario evaluasi yang berbeda. Penelitian sebelumnya tidak memisahkan data validasi dan data pengujian secara independen, sedangkan penelitian ini menerapkan evaluasi yang lebih ketat untuk memperoleh gambaran kemampuan generalisasi model yang lebih realistis.

Temuan ini menunjukkan bahwa peningkatan performa model tidak hanya bergantung pada kualitas data hasil augmentasi, tetapi juga dipengaruhi oleh kemampuan model dalam menangani ketidakseimbangan data dan mempertahankan representasi fitur yang relevan. Dengan demikian, penerapan *Global Max Pooling*, *class weighting*, dan evaluasi independen memberikan kontribusi penting dalam pengembangan sistem *Automatic Essay Scoring* berbasis bahasa Indonesia yang lebih stabil dan dapat diandalkan.

4. KESIMPULAN DAN SARAN

Penelitian ini mengusulkan optimasi model *Automatic Essay Scoring* (AES) berbasis BiLSTM-FastText pada data hasil augmentasi semantik EDA_Sinonim_IndoBERT menggunakan dataset UKARA. Optimasi dilakukan melalui penerapan *Global Max Pooling* untuk meningkatkan representasi fitur serta *class weighting* untuk mengatasi ketidakseimbangan data. Hasil eksperimen menunjukkan bahwa model mampu mencapai akurasi sebesar 93,49% pada data validasi dan 78,00% pada data pengujian independen. Temuan ini menunjukkan bahwa augmentasi semantik efektif dalam meningkatkan proses pembelajaran model, namun kemampuan generalisasi terhadap data yang belum pernah dilihat masih menjadi tantangan. Selain itu, *class weighting* terbukti membantu meningkatkan kemampuan model dalam mengenali kelas minoritas sehingga menghasilkan performa klasifikasi yang lebih seimbang. Secara keseluruhan, hasil penelitian menegaskan bahwa kualitas sistem AES tidak hanya ditentukan oleh metode augmentasi, tetapi juga oleh strategi optimasi model dan pelatihan yang digunakan.

Penelitian selanjutnya dapat difokuskan pada pengembangan mekanisme penyaringan (*filtering*) terhadap data hasil augmentasi untuk mengurangi *noise* semantik yang berpotensi mengubah makna jawaban siswa. Selain itu, pendekatan berbasis *attention mechanism*, *transformer*, maupun *contextual embedding* yang lebih mutakhir dapat dieksplorasi guna meningkatkan kemampuan model dalam memahami konteks jawaban secara lebih mendalam. Evaluasi yang lebih komprehensif melalui *stratified k-fold cross validation* serta pengujian pada dataset dengan karakteristik dan tingkat kompleksitas yang lebih beragam juga perlu dilakukan untuk menghasilkan sistem *Automatic Essay Scoring* (AES) yang lebih stabil, adaptif, dan representatif terhadap kondisi pembelajaran yang sesungguhnya.

REFERENSI

Al-Sultan, A., dkk. (2025). A novel approach for mitigating class imbalance in text classification. *IEEE Access*, *PP*(99), 1-1. https://doi.org/10.1109/ACCESS.2025.3611636

Aliyah, N. E., Sholikah, R. W., Firdausi, H., Ciptaningtyas, H. T., & Sabilla, I. A. (2025). Enhancing Automated Essay Scoring in Bahasa Indonesia with IndoBERT and IndoSBERT. *2025*

International Conference on Smart Computing, IoT and Machine Learning (SIML). IEEE. https://doi.org/10.1109/SIML65326.2025.11080721

Ariyus, D., Manongga, D., & Sembiring, I. (2024). Enhancing Sentiment Analysis of Indonesian Tourism Video Content Commentary on TikTok: A FastText and Bi-LSTM Approach. *Engineering, Technology & Applied Science Research*, *14*(6), 18020-18028. https://doi.org/10.48084/etasr.8859

Dhini, B. F., Girsang, A. S., Sufandi, U. U., & Kurniawati, H. (2023). Automatic essay scoring for discussion forum in online learning. *Asian Association of Open Universities Journal*, *18*(3), 262-279. https://doi.org/10.1108/AAOUJ-05-2023-0050

Fadilah, N., & Priyanta, S. (2022). Automatic Essay Scoring Using Data Augmentation in Bahasa Indonesia. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, *16*(4), 401. https://doi.org/10.22146/ijccs.76396

Fadilah, N., & Zain, S. G. (2024). Rancang Bangun Sistem Penilaian Tes Essai Berbasis Web Di Testing Center UNM. PISCES: Journal of Progressive Information, Security, Computer and Embedded System, 2(1), 38–45. https://journal.diginus.id/PISCES/article/view/296

Mohammad (2025). *CoGate-LSTM: A lightweight recurrent model that addresses extreme class imbalance through cosine-similarity feature gating*. arXiv preprint arXiv:2510.17018v2. https://arxiv.org/html/2510.17018v2

Nur Azizah, A., Falach Asy'ari, M., Wisma Dwi Prastya, I., & Purwitasari, D. (2023). Easy Data Augmentation untuk Data yang Imbalance pada Konsultasi Kesehatan Daring. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 10(5), 1095–1104. https://doi.org/10.25126/jtiik.20231057082

Pratama, M. A., & Budi, I. (2026). Improved Text Classification for Indonesian Hate Speech Detection: FastText-LSTM Model with Easy Data Augmentation. *Jurnal Sistem Informasi (JSI)*, 12(1), 45–56. https://doi.org/10.25126/jsi.20261219637

Rahma, I. A., & Suadaa, L. H. (2023). Penerapan Text Augmentation untuk Mengatasi Data yang Tidak Seimbang pada Klasifikasi Teks Berbahasa Indonesia. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 10(6), 1329–1340. https://doi.org/10.25126/jtiik.2023107325

Ramadhan, T. A., & Purwarianti, A. (2025). Perbandingan Kinerja Model BiLSTM dan IndoBERT dalam Deteksi Karakteristik Teks Media Online di Indonesia. *Jurnal Informatika dan Teknik Elektro Terapan (JITET)*, 13(2), 201–212. https://journal.eng.unila.ac.id/index.php/jitet/article/view/9365

Reknadi, D., Rohman, M. G., Mustain, M., & Utomo, A. F. L. (2025). Adaptation of Contrastive Learning and Augmentation for Indonesian Product Review Classification on Unbalanced Data Using Deep Learning and NLP. *Generation Journal*, 9(2), 115–126. https://ojs.unpkediri.ac.id/index.php/gj/article/view/22730

Setiawan, H., & Prasetyo, E. (2024). Fine-Tuned IndoBERT Based Model and Data Augmentation for Indonesian Language Paraphrase Identification. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 8(3), 415–423. https://doi.org/10.29207/resti.v8i3.5122