

Sistem Penilaian Esai Otomatis Berbasis Web Menggunakan Model BERT dan *Levenshtein Distance*

Aisya Febrianti Azzaroh¹, Arda Surya Editya², Ahmad Khoir Al-Haq³

¹Informatika, fakultas Ilmu Komputer, Universitas Nahdlatul Ulama Sidoarjo

²Informatika, fakultas Ilmu Komputer, Universitas Nahdlatul Ulama Sidoarjo

³Informatika, fakultas Ilmu Komputer, Universitas Nahdlatul Ulama Sidoarjo

¹123421009.mhs@unusida.ac.id. ² ardasurya.tif@unusida.ac.id. ³ ahmadkhoiralhaq@unusida.ac.i

Abstract

Manual essay grading systems have limitations of time consuming efficiency and objectivity. To overcome this matter, this study developed a web-based automatic grading system using the BERT and Levenshtein Distance models. This system combines semantic grading with the BERT model (80%) and syntactic grading using Levenshtein Distance (20%). This platform is built using the Flask framework and provides two main views: first, the student view to answer questions then get results directly. Second, the teacher view to manage questions, answer keys, student data, assessment results, grade histograms and system accuracy. The test results show that the system is able to provide assessments that are close to manual assessments with fairly high accuracy. This system is expected to help teachers in assessing efficiently and consistently.

Keywords: Automatic Assessment, BERT, Levenshtein Distance, Web, Flask, Essay, Learning Abstract

Abstrak

Sistem penilaian esai secara manual memiliki keterbatasan dalam hal efisiensi waktu dan objektivitas. Untuk mengatasi hal tersebut, penelitian ini mengembangkan sistem penilaian otomatis berbasis web menggunakan model BERT dan *Levenshtein Distance*. Sistem ini menggabungkan penilaian semantik dengan model BERT (80%) dan penilaian sintaksis menggunakan *Levenshtein Distance* (20%). Platform ini dibangun menggunakan framework Flask dan menyediakan dua tampilan utama: tampilan siswa untuk menjawab soal dan memperoleh hasil secara langsung, serta tampilan guru untuk mengelola soal, kunci jawaban, data siswa, hasil penilaian, histogram nilai, dan akurasi sistem. Hasil pengujian menunjukkan bahwa sistem mampu memberikan penilaian yang mendekati penilaian manual dengan akurasi yang cukup tinggi. Sistem ini diharapkan dapat membantu guru dalam menilai secara efisien dan konsisten.

Kata kunci: Penilaian Otomatis, BERT, *Levenshtein Distance*, Web, Flask, Esai, Pembelajaran

© 2025 Jurnal Pustaka AI

1. Pendahuluan

Evaluasi adalah proses menentukan tingkat keberhasilan dalam proses pengajaran. evaluasi pembelajaran biasanya terdiri dari tes pilihan ganda dan tes uraian atau esai[1]. Di era digital saat ini, teknologi telah menjadi bagian integral dalam berbagai aspek kehidupan, termasuk pendidikan. Salah satu tantangan terbesar dalam dunia pendidikan adalah bagaimana melakukan penilaian secara efisien dan objektif, terutama dalam konteks penilaian soal esai yang membutuhkan analisis mendalam terhadap jawaban siswa. Dalam literasi baca tulis, yang merupakan bagian penting dari pendidikan dasar, tantangan ini semakin terasa.

Berdasarkan data dari Indeks Pembangunan Literasi Masyarakat (IPLM) nasional tahun 2024 dengan skor 73,52, melampaui target 71,4 dan hasil tahun sebelumnya yang berada di angka 69,42 tantangan dalam literasi anak-anak Indonesia masih signifikan[2]. Sebagai upaya untuk meningkatkan minat literasi di kalangan siswa, Kemendikbud telah mendistribusikan 4,6 juta buku ke sekolah-sekolah di seluruh Indonesia[3]. Penurunan skor ini menunjukkan perlunya intervensi lebih lanjut untuk meningkatkan keterampilan literasi siswa di berbagai tingkatan Pendidikan.

Kendati teknologi telah diadopsi dalam berbagai aspek pendidikan, seperti pengelolaan kelas dan penilaian soal pilihan ganda, otomatisasi penilaian soal esai masih menjadi tantangan besar[4]. Kebanyakan sistem penilaian otomatis hanya mampu menilai jawaban terstruktur seperti pilihan ganda, sementara penilaian esai masih membutuhkan intervensi manual[5]. Hal ini menciptakan kesenjangan karena proses manual tidak hanya memakan waktu tetapi juga rentan terhadap subjektivitas. Oleh karena itu, diperlukan pendekatan yang inovatif untuk mengisi gap tersebut dengan solusi yang dapat menilai esai secara objektif, akurat, dan efisien.

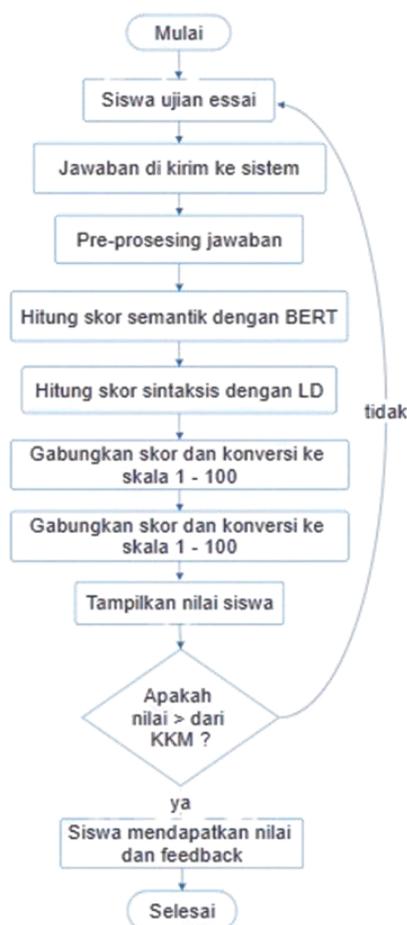
Untuk menjawab tantangan dalam penilaian esai, penelitian ini mengusulkan penerapan metode Levenshtein Distance dan BERT. Levenshtein Distance digunakan untuk mengukur kemiripan teks secara sintaksis dengan menghitung jumlah minimum perubahan yang dibutuhkan untuk mengubah satu string menjadi string lainnya. Sementara itu, Bidirectional Encoder Representations from Transformers (BERT) merupakan salah satu model Natural Language Processing (NLP) berbasis deep learning yang digunakan untuk menilai kesamaan makna (semantik) antar teks[6]. Dengan menggabungkan kedua metode ini dalam sistem penilaian otomatis, diharapkan proses evaluasi menjadi lebih cepat, objektif, dan andal.

2. Metode Penelitian

Dalam tahap pertama rancangan sistem, dikembangkan flowchart sebagai kerangka kerja utama yang memetakan alur dari input esai siswa hingga keluaran skor akhir[7]. Proses bermula ketika siswa mengumpulkan esainya melalui antarmuka web, kemudian teks esai tersebut diproses melalui tahap *preprocessing* untuk membersihkan dan menormalisasi teks (misalnya *tokenisasi*, *lowercasing*, dan penghilangan tanda baca)[8]. Setelah itu, model BERT yang telah dilatih khusus digunakan untuk menghitung skor semantik, yaitu seberapa dekat esai siswa dengan jawaban ideal dari segi makna konteks.

Metode ini telah terbukti efektif dalam konteks otomatisasi penilaian teks berstruktur, seperti penelitian ini mengenai penilaian terjemahan otomatis yang menyatakan bahwa penggunaan BERT-based scoring system mampu mencapai reliabilitas tinggi ($\geq 0,9$) dalam menentukan kualitas teks .

Berikut gambar flowchart yang ada di sistem penilaian otomatis soal essai :



Gambar 1. Flowchart sistem

Tahap berikutnya melibatkan evaluasi kesamaan sintaksis antara teks siswa dan jawaban kunci menggunakan Levenshtein Distance, yaitu algoritme pengukuran jarak berbasis minimum operasi (*insertions, deletions, substitutions*)[9]. Skor semantik dari BERT dan skor sintaksis dari Levenshtein Distance, selanjutnya dikonversi ke rentang nilai 0–100 agar setara, lalu digabungkan dengan bobot 80% semantik dan 20% sintaksis[10]. Kombinasi ini dirancang agar sistem mampu menilai esai secara komprehensif, mempertimbangkan kedalaman makna sekaligus struktur kalimat.

Hasil yang terintegrasi kemudian dibandingkan dengan ambang batas Kriteria Ketuntasan Minimal (KKM), sehingga dapat memberikan umpan balik otomatis yang objektif dan adaptif kepada siswa melalui platform berbasis web.

2.1. Pengumpulan Data

Pengumpulan data dalam penelitian ini dilakukan di SDN Cangkring 1 dengan bekerja sama langsung dengan guru wali kelas V. Peneliti meminta data berupa soal, kunci jawaban, dan jawaban siswa untuk beberapa jenis evaluasi, yaitu ulangan harian, Ujian Tengah Semester (UTS), dan Ujian Akhir Semester (UAS) pada mata pelajaran Bahasa Indonesia dan Pendidikan Kewarganegaraan (PKn). Data ini dikumpulkan untuk membentuk dataset yang akan digunakan sebagai acuan dalam melatih dan menguji sistem penilaian otomatis. Setiap esai siswa yang dikumpulkan telah dinilai terlebih dahulu secara manual oleh guru, sehingga dapat digunakan sebagai pembandingan atau ground truth untuk evaluasi sistem.

Secara keseluruhan, terkumpul sebanyak 100 esai yang terdiri dari berbagai variasi soal dan jawaban siswa. Semua data tersebut dikodekan secara anonim guna menjaga kerahasiaan identitas siswa. Penggunaan data nyata dari sekolah dasar ini bertujuan untuk memastikan bahwa sistem penilaian yang dikembangkan mampu bekerja secara optimal dalam konteks pembelajaran di tingkat sekolah dasar. Pendekatan ini sejalan dengan praktik umum dalam pengembangan sistem penilaian esai otomatis yang menggunakan data nyata dan penilaian manusia sebagai acuan utama dalam validasi sistem. Hal ini juga sesuai dengan hasil penelitian sebelumnya yang menekankan pentingnya keterlibatan data nyata dalam proses pelatihan dan pengujian sistem penilaian otomatis.

2.2. Metode Penggabungan Skor

Dalam sistem penilaian esai otomatis yang dikembangkan dalam penelitian ini, proses penilaian dilakukan terhadap 10 butir soal esai dengan pendekatan gabungan antara penilaian semantik dan sintaksis. Setiap jawaban siswa dinilai dari dua sisi: makna (semantik) dan struktur penulisan (sintaksis). Skor semantik dihitung menggunakan model BERT yang mengukur kedekatan makna antara jawaban siswa dan kunci jawaban, sedangkan skor sintaksis dihitung menggunakan algoritma Levenshtein Distance yang mengukur kemiripan bentuk teks berdasarkan jumlah operasi edit yang diperlukan untuk menyamakan kedua teks. Tujuan penggabungan ini adalah untuk memastikan bahwa sistem tidak hanya menilai kata demi kata secara literal, tetapi juga memahami isi dan konteks dari jawaban yang diberikan siswa.

Setiap jawaban pada satu soal akan diberi skor akhir dengan bobot 80% untuk aspek semantik dan 20% untuk aspek sintaksis. Skor total untuk seluruh soal dihitung dengan menjumlahkan skor dari 10 soal, kemudian dikonversi ke dalam skala 0–100. Formula yang digunakan untuk menghitung skor akhir tiap soal adalah:

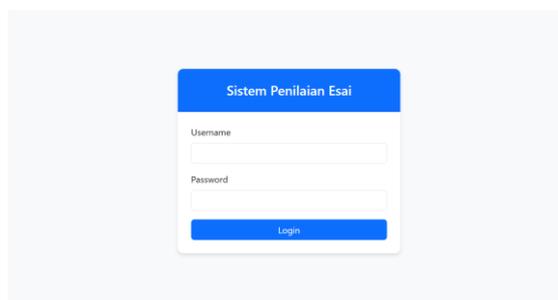
$$\text{Skor soal} = (0.8 \times \text{skor semantik}) + (0.2 \times \text{skor sintaksis})$$

Selanjutnya, nilai akhir siswa dihitung dengan menjumlahkan skor dari seluruh soal dan membaginya agar sesuai dengan skala 0 sampai 100. Pendekatan ini memungkinkan sistem untuk memberikan penilaian yang lebih adil dan komprehensif, mengingat siswa bisa saja menyampaikan jawaban dengan makna yang tepat namun menggunakan kata-kata atau struktur yang berbeda dari kunci jawaban. Dengan demikian, sistem ini tidak hanya menilai berdasarkan kemiripan bentuk tulisan, tetapi juga menilai sejauh mana pemahaman siswa terhadap isi pertanyaan.

3. Hasil dan Pembahasan

3.1. Implementasi Tampilan Antar Muka

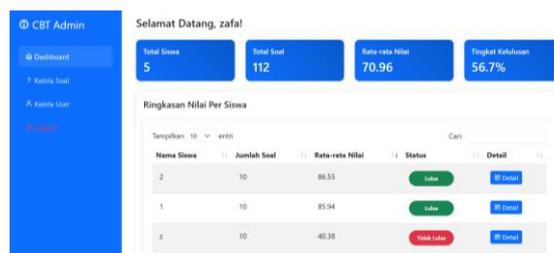
Desain awal antarmuka bagian login menggunakan elemen-elemen sederhana dengan warna putih dan biru seperti gambar di bawah ini.



Gambar 2. Tampilan Login

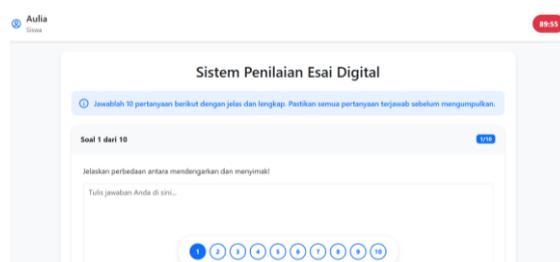
Tampilan login pada sistem ini digunakan sebagai pintu masuk utama bagi pengguna. Siswa dapat melakukan login setelah sebelumnya didaftarkan oleh guru melalui halaman admin. Sementara itu, guru yang sudah memiliki akun juga bisa langsung masuk ke sistem. Form login terdiri dari kolom email dan password, serta tombol masuk, yang dirancang sederhana agar mudah digunakan oleh semua pengguna.

Selanjutnya, guru yang telah terdaftar sebagai admin dapat langsung melakukan login. Setelah berhasil masuk, sistem akan menampilkan tampilan dashboard admin seperti yang diperlihatkan pada gambar di bawah ini.



Gambar 3. Tampilan Dashboard Guru

Pada Gambar 3, tampilan dashboard guru menampilkan sidebar yang berisi tombol navigasi seperti Kelola Soal dan Kelola Users. Melalui menu Kelola Soal, guru dapat menambahkan soal dengan dua cara, yaitu mengimpor soal dari file Excel atau CSV, serta menambahkan soal secara manual. Guru juga dapat melihat detail jawaban siswa beserta kunci jawaban yang digunakan sebagai acuan penilaian.



Gambar 4. Tampilan Ujian

Pada halaman ini, siswa dapat mulai mengerjakan soal yang tersedia satu per satu. Soal akan ditampilkan secara acak dari database, dengan total sebanyak 10 soal yang ditampilkan. Siswa dapat langsung mengerjakan karena sistem telah dilengkapi dengan batas waktu pengerjaan, yaitu selama 90 menit. Waktu ini ditampilkan secara real-time agar siswa dapat memantau sisa waktu ujian.

3.2. Proses Penilaian pada Sistem

Pada tahap ini ketika seluruh jawaban siswa terkumpul, sistem secara otomatis menjalankan proses penilaian. Proses ini diawali dengan pengukuran kesamaan semantik menggunakan model BERT, diikuti oleh penghitungan kesamaan sintaksis menggunakan algoritma Levenshtein Distance. Hasil dari kedua proses ini berupa dua jenis skor: skor semantik dan skor sintaksis, yang masing-masing merepresentasikan aspek makna dan struktur kalimat dari jawaban siswa.

Kedua skor tersebut kemudian digabungkan dengan pembobotan tertentu, yang mana perhitungan skor sintaksis dalam pembobotan ini merujuk pada Persamaan (1).

Nilai akhir yang dihasilkan dikonversi ke dalam skala 0–100 dan ditampilkan kepada guru serta siswa. Pendekatan ini memungkinkan penilaian yang lebih komprehensif dan objektif. Gambar berikut menampilkan antarmuka sistem yang memperlihatkan hasil penilaian otomatis berupa skor semantik, skor sintaksis, skor akhir, dan status kelulusan siswa seperti gambar di bawah ini.

Menampilkan 1 sampai 6 dari 6 entri

Detail Jawaban Siswa

Nama Siswa	Soal & Jawaban	Skor Semantik	Skor Sintaksis	Skor Akhir	Status
1	Klik untuk detail	85.7%	85.7%	85.7%	Tidak Lulus
1	Klik untuk detail	86.1%	86.1%	86.1%	Lulus
1	Klik untuk detail	100.0%	100.0%	100.0%	Lulus
1	Klik untuk detail	85.7%	85.7%	85.7%	Tidak Lulus
1	Klik untuk detail	100.0%	100.0%	100.0%	Lulus

Gambar 5. Tampilan Detail Nilai

3.3. Kelebihan dan Kekurangan Sistem

Sistem penilaian esai otomatis yang dikembangkan memiliki sejumlah kelebihan yang dapat memberikan dampak positif bagi proses pembelajaran. Salah satu kelebihannya adalah efisiensi waktu dalam proses penilaian. Dengan

integrasi model BERT untuk analisis semantik dan algoritma Levenshtein Distance untuk analisis sintaksis, sistem mampu memberikan hasil penilaian secara instan setelah siswa menyelesaikan ujian. Hal ini tentu sangat membantu guru dalam mempercepat proses koreksi, terutama ketika jumlah siswa dan soal yang dinilai cukup banyak.

Selain itu, sistem ini juga mampu menghasilkan penilaian yang lebih objektif dan konsisten. Dengan bobot yang ditentukan untuk skor semantik dan sintaksis, penilaian tidak lagi bergantung pada interpretasi subjektif guru. Pendekatan ini memungkinkan siswa mendapatkan penilaian yang adil, serta dapat dijadikan sebagai sarana umpan balik yang cepat dan tepat untuk meningkatkan kualitas pembelajaran, khususnya dalam mata pelajaran yang menekankan kemampuan literasi dan argumentasi, seperti Bahasa Indonesia dan Pendidikan Kewarganegaraan (PKn).

Namun, sistem ini juga memiliki beberapa keterbatasan. Salah satunya adalah tidak cocok untuk digunakan pada mata pelajaran yang bersifat numerik atau eksak, seperti Matematika atau IPA, karena bentuk soal dan kriteria penilaiannya sangat berbeda dengan soal esai literasi. Sistem ini dirancang khusus untuk mengevaluasi pemahaman makna dan kemiripan penulisan dalam jawaban berbasis teks, sehingga tidak relevan untuk menilai jawaban berupa rumus atau perhitungan angka.

Selain itu, sistem masih terbatas pada data pelatihan dari dua mata pelajaran, yaitu Bahasa Indonesia dan PKn. Hal ini membuat kemampuan generalisasi sistem terhadap soal dari mata pelajaran lain menjadi kurang optimal. Sistem belum sepenuhnya mampu mengenali konteks atau variasi bahasa dari mata pelajaran dengan karakteristik yang berbeda, sehingga perlu dilakukan pengembangan lebih lanjut dengan melibatkan dataset dari berbagai bidang studi agar cakupan dan akurasi sistem semakin luas dan stabil.

3.4 Manfaat Sistem

Sistem penilaian esai otomatis yang dikembangkan memberikan sejumlah manfaat penting dalam mendukung proses evaluasi pembelajaran, khususnya pada mata pelajaran yang berbasis literasi seperti Bahasa Indonesia dan PKn. Dengan memanfaatkan kecerdasan buatan, sistem ini mampu mempercepat proses penilaian, mengurangi beban kerja guru, serta menghasilkan penilaian yang lebih objektif dan konsisten.

Selain itu, siswa dapat memperoleh umpan balik secara langsung setelah menyelesaikan ujian, sehingga proses pembelajaran menjadi lebih efektif dan responsif. Sistem ini juga berpotensi untuk diintegrasikan ke dalam platform pembelajaran digital yang lebih luas guna menunjang digitalisasi pendidikan di era modern.

4. Kesimpulan

Penelitian ini menghasilkan sebuah sistem penilaian esai otomatis berbasis web yang menggabungkan model BERT untuk analisis semantik dan algoritma Levenshtein Distance untuk analisis sintaksis. Sistem ini mampu memberikan penilaian secara cepat, objektif, dan konsisten terhadap soal esai berbasis literasi, khususnya pada mata pelajaran Bahasa Indonesia dan Pendidikan Kewarganegaraan (PKn). Hasil evaluasi menunjukkan bahwa kombinasi pendekatan semantik dan sintaksis mampu merepresentasikan kemampuan siswa secara lebih menyeluruh dibandingkan penilaian manual yang rawan subjektivitas.

Namun demikian, sistem ini masih memiliki keterbatasan, terutama untuk soal pada mata pelajaran non-literasi seperti Matematika. Pengembangan lebih lanjut diperlukan dengan memperluas dataset dari berbagai mata pelajaran dan melakukan uji coba pada skala yang lebih besar untuk meningkatkan akurasi dan cakupan sistem secara keseluruhan.

Daftar Rujukan

- [1] N. L. Kinanti and A. Qoiriah, "Sistem Penilaian Otomatis Jawaban Esai Bahasa Indonesia Berdasarkan Kemiripan Kalimat Menggunakan Syntactic-Semantic Similarity," *JINACS (Journal Informatics Comput. Sci.*, vol. 2, no. 2, pp. 136–143, 2020.
- [2] A. Khoirun, "IPLM 2024 Catat Rekor Tinggi, Literasi Nasional Semakin Meningkat," www.perpusnas.go.id. Accessed: Mar. 19, 2025.
- [3] S. Putrawangsa and U. Hasanah, "Integrasi Teknologi Digital Dalam Pembelajaran Di Era Industri 4.0," *J. Tatsqif*, vol. 16, no. 1, pp. 42–54, 2018, doi: 10.20414/jtq.v16i1.203.
- [4] M. Shermis and J. Burstein, *Handbook of Automated Essay Evaluation: Current Applications and New Directions*, 1st ed. New York: Routledge, 2013.
- [5] N. Thompson, "What is Automated Essay Scoring?," *Assess.com*, Apr. 25, 2023.
- [6] Chandradev, V., Suarjaya, I. M. A. D., & Bayupati, I. P. A. (2023). Analisis sentimen review hotel menggunakan metode deep learning bert. *Jurnal Buana Informatika*, 14(02), 107-116.
- [7] Lim, C. T., Bong, C. H., Wong, W. S., & Lee, N. K. (2021). A comprehensive review of Automated Essay Scoring (AES): Research and development. *Pertanika Journal of Science & Technology*, 29(3), 1875–1899.
- [8] V. No, F. T. Sabilillah, S. Winarno, and R. B. Abiyi, "Edumatic : Jurnal Pendidikan Informatika Implementasi BERT dan Cosine Similarity untuk Rekomendasi Dosen Pembimbing berdasarkan Judul Tugas Akhir," vol. 8, no. 2, pp. 585–594, 2024, doi: 10.29408/edumatic.v8i2.27791

- [9] Sugiarto, I. G. S. M. Diyasa, and I. N. Diana, "Levenshtein distance algorithm analysis on enrollment and disposition of letters application," Proceeding - 6th Inf. Technol. Int. Semin. ITIS 2020, vol. 1, pp. 198–202, 2020, doi: 10.1109/ITIS50118.2020.9321030.
- [10] Syamsiah, S. (2019). Perancangan Flowchart dan Pseudocode Pembelajaran Mengenal Angka dengan Animasi untuk Anak PAUD Rambutan. *STRING (Satuan Tulisan Riset Dan Inovasi Teknologi)*, 4(1), 86-93.
-