



SISTEM COMPUTER BASED TEST DENGAN PENILAIAN OTOMATIS SOAL ESAI MENGGUNAKAN ALGORITMA TEXT SIMILARITY BERBASIS AI

Anggit Hardiyanto¹, Moh Muhtarom², Joni Maulindar³

^{1,2,3}*Universitas Duta Bangsa Surakarta, Surakarta, Indonesia*

Penulis Korespondensi: anggithardiyanto228@gmail.com

ABSTRAK

Transformasi digital di bidang pendidikan terus berkembang, salah satunya melalui Computer Based Test (CBT) sebagai sarana evaluasi pembelajaran. Meskipun CBT efektif untuk soal objektif, penilaian esai masih menjadi kendala karena memerlukan waktu lama dan rawan subjektivitas. Penelitian ini bertujuan membangun sistem CBT dengan penilaian esai otomatis berbasis algoritma Text Similarity dan kecerdasan buatan, guna meningkatkan efisiensi, konsistensi, dan objektivitas evaluasi. Pengembangan sistem menggunakan metode Waterfall, meliputi analisis kebutuhan, perancangan, implementasi, pengujian, dan pemeliharaan. Jawaban siswa dan kunci jawaban diproses melalui tahapan Natural Language Processing (NLP), yaitu case folding, tokenizing, stopword removal, dan stemming. Selanjutnya, bobot kata dihitung menggunakan TF-IDF, dan tingkat kemiripan jawaban dihitung dengan algoritma Cosine Similarity untuk menghasilkan nilai otomatis. Hasil penelitian menunjukkan sistem mampu mengintegrasikan ujian berbasis komputer dan penilaian esai otomatis dalam satu platform. Pada salah satu skenario, Cosine Similarity menghasilkan kemiripan 74%, menunjukkan kesesuaian tinggi antara jawaban siswa dan kunci jawaban. Pengujian Black Box menunjukkan seluruh fitur utama berjalan sesuai kebutuhan dengan keberhasilan 100%. Sistem ini memberikan hasil penilaian secara cepat dan konsisten, serta mendukung digitalisasi pendidikan. Penelitian selanjutnya disarankan mengembangkan model berbasis transformer untuk meningkatkan akurasi penilaian.

Kata Kunci: *Computer Based Test, Penilaian Esai Otomatis, Natural Language Processing, Cosine Similarity*

Riwayat Artikel :

Tanggal diterima : 15-06-2026

Tanggal terbit : 23-07-2026

Kutipan :

Hardiyanto, A., Muhtarom, M., & Maulindar, J. (2026). SISTEM COMPUTER BASED TEST DENGAN PENILAIAN OTOMATIS SOAL ESAI MENGGUNAKAN ALGORITMA TEXT SIMILARITY BERBASIS AI. INFOTECH Journal, 12(2), 164–171. <https://doi.org/10.31949/infotech.v12i2.18754>



1. PENDAHULUAN

1.1. Latar Belakang

Kemajuan teknologi informasi telah menggerakkan perubahan digital di berbagai bidang, tak terkecuali dunia pendidikan. Salah satu wujud nyata dari perkembangan tersebut adalah penerapan Computer Based Test (CBT), yang memungkinkan proses penilaian hasil belajar dilaksanakan secara digital dan terpadu dalam satu sistem. Penerapan CBT memberikan berbagai keuntungan, seperti mempercepat pelaksanaan ujian, memudahkan pengelolaan data, mengurangi penggunaan kertas, serta mempercepat proses pengolahan hasil evaluasi. Oleh karena itu, sistem CBT semakin banyak diterapkan pada berbagai institusi pendidikan sebagai bagian dari upaya digitalisasi proses pembelajaran (Ulum & Hidayat, 2024).

Namun demikian, sebagian besar sistem CBT yang diterapkan hingga saat ini masih menitikberatkan pada penilaian soal bertipe objektif seperti pilihan ganda. Penilaian soal esai masih banyak dilakukan secara manual karena memerlukan interpretasi terhadap jawaban peserta. Kondisi ini menyebabkan proses koreksi membutuhkan waktu yang relatif lama, terutama ketika jumlah peserta ujian cukup banyak. Selain itu, penilaian manual juga berpotensi menimbulkan subjektivitas dan inkonsistensi karena adanya perbedaan persepsi antar penilai (Pradani & Suadaa, 2023; Rokhman et al., 2025).

Perkembangan teknologi Artificial Intelligence (AI) membuka peluang untuk menjawab tantangan tersebut melalui penerapan sistem penilaian esai secara otomatis. Salah satu pendekatan yang umum diterapkan dalam hal ini adalah metode text similarity, yakni teknik yang bekerja dengan mengukur sejauh mana kemiripan antara jawaban peserta ujian dan jawaban acuan yang telah ditetapkan. Sejumlah penelitian mengungkapkan bahwa metode ini mampu mempercepat proses penilaian esai sekaligus menghasilkan penilaian yang lebih objektif dibandingkan dengan koreksi yang dilakukan secara manual (Abdurrahman et al., 2024; Pane et al., 2024). Akan tetapi, sebagian besar penelitian terdahulu masih terpusat pada pengembangan algoritma penilaian atau pengujian tingkat akurasi metode secara terpisah dan belum banyak yang menggabungkan teknologi tersebut ke dalam sistem CBT yang dapat dimanfaatkan secara langsung dalam kegiatan evaluasi pembelajaran. Berdasarkan kondisi tersebut, diperlukan pengembangan sistem CBT yang mampu melakukan penilaian soal esai secara otomatis dengan memanfaatkan teknologi AI. Penelitian ini bertujuan mengembangkan Sistem Computer Based Test dengan Penilaian Otomatis Soal Esai Menggunakan Algoritma Text Similarity Berbasis AI dengan memanfaatkan Natural Language Processing (NLP), metode pembobotan TF-IDF, serta algoritma Cosine Similarity. Sistem yang dibangun diharapkan dapat memberikan kontribusi dalam meningkatkan efisiensi pelaksanaan proses evaluasi, mengurangi subjektivitas penilaian, serta memberikan hasil

penilaian yang lebih cepat, objektif, dan informatif bagi guru maupun peserta didik.

1.2. Tinjauan Pustaka

1) Computer Based Test (CBT)

Computer Based Test (CBT) merupakan sistem pelaksanaan ujian yang memanfaatkan komputer sebagai media utama dalam menyajikan soal, pelaksanaan pengerjaan ujian, serta pengolahan hasil evaluasi secara digital. CBT dikembangkan untuk meningkatkan efisiensi proses evaluasi pembelajaran melalui otomatisasi pengelolaan soal, penyimpanan data, dan penyajian hasil ujian. Penggunaan CBT mampu mengurangi penggunaan kertas, mempercepat proses evaluasi, serta meningkatkan keamanan data ujian. Selain itu, CBT memungkinkan integrasi berbagai teknologi pendukung yang dapat meningkatkan kualitas proses penilaian (Abdurrahman et al., 2024).

Dalam penelitian ini, CBT tidak hanya berperan sebagai media penyelenggaraan ujian semata, tetapi juga berfungsi sebagai platform yang memadukan proses penilaian otomatis untuk soal esai. Dengan demikian, seluruh proses evaluasi dapat dijalankan secara lebih efisien, objektif, dan konsisten dibandingkan metode konvensional.

2) Penilaian Esai Otomatis (Automated Essay Scoring)

Penilaian esai otomatis (Automated Essay Scoring/AES) merupakan pendekatan evaluasi yang memanfaatkan teknologi kecerdasan buatan untuk melakukan penilaian terhadap jawaban esai tanpa campur tangan manusia secara langsung. Sistem AES bekerja dengan menganalisis isi jawaban siswa dan membandingkannya dengan jawaban referensi menggunakan teknik pemrosesan bahasa alami serta algoritma pengukuran kemiripan teks (Pradani & Suadaa, 2023; Rokhman et al., 2025).

Penerapan AES mampu meningkatkan konsistensi penilaian karena sistem melakukan evaluasi berdasarkan aturan dan model yang sama terhadap seluruh jawaban peserta (Pradani & Suadaa, 2023). Dalam penelitian ini, AES diimplementasikan sebagai fitur utama dalam sistem CBT sehingga hasil penilaian dapat diperoleh secara otomatis setelah peserta menyelesaikan ujian.

3) Natural Language Processing (NLP)

Natural Language Processing (NLP) adalah salah satu bidang dalam ilmu kecerdasan buatan yang berfokus pada kemampuan komputer untuk memahami, mengolah, dan menganalisis bahasa yang digunakan manusia, baik yang berupa teks maupun tuturan lisan. Melalui NLP, sistem komputer dapat mengekstrak informasi dari dokumen berbasis teks sehingga dapat dimanfaatkan untuk berbagai keperluan, seperti klasifikasi teks, analisis sentimen, maupun penilaian secara otomatis (Maulidya Prastita Syah et al., 2025; Sihombing, 2022). Dalam penelitian ini, NLP diterapkan sebagai landasan dalam memproses jawaban esai melalui sejumlah tahapan preprocessing, antara lain:

- a. Case Folding, yaitu proses mengubah seluruh karakter teks menjadi huruf kecil secara seragam.
- b. Tokenizing, yaitu proses membagi teks kalimat menjadi unit-unit kata yang lebih kecil.
- c. Stopword Removal, yaitu proses menyaring dan membuang kata-kata yang tidak memiliki makna berarti.
- d. Stemming, yaitu proses mereduksi kata berimbuhan menjadi bentuk dasarnya.

Keseluruhan rangkaian tahapan tersebut dimaksudkan untuk menghasilkan representasi teks yang lebih bersih dan terorganisir, sehingga proses penghitungan tingkat kemiripan antar teks dapat berjalan secara lebih optimal (Sihombing, 2022).

4) Algoritma Text Similarity

Text Similarity adalah sebuah metode yang dimanfaatkan untuk mengukur sejauh mana tingkat kesesuaian antara dua dokumen atau teks berdasarkan karakteristik atau fitur tertentu yang dimilikinya. Dalam konteks penilaian esai secara otomatis, text similarity dimanfaatkan untuk membandingkan jawaban yang disampaikan oleh peserta ujian dengan jawaban referensi yang sebelumnya telah ditentukan oleh pengajar (Pane et al., 2024; Rokhman et al., 2025).

Semakin besar nilai kemiripan yang dihasilkan, maka semakin erat kedekatan makna antara jawaban peserta dan kunci jawaban yang telah ditentukan sebelumnya. Dengan demikian, metode ini menjadi elemen krusial dalam sistem penilaian otomatis karena mampu menghasilkan skor penilaian berdasarkan tingkat kesesuaian isi jawaban peserta (Pane et al., 2024).

5) Cosine Similarity

Cosine Similarity merupakan suatu algoritma yang diterapkan untuk menghitung tingkat kemiripan antara dua buah dokumen dengan cara menghitung sudut yang terbentuk dari representasi vektor kedua dokumen tersebut. Nilai keluaran yang dihasilkan berada pada rentang 0 hingga 1, di mana semakin besar nilai yang mendekati angka 1 menunjukkan semakin erat tingkat kemiripan antara kedua dokumen yang diperbandingkan (Abdurrahman et al., 2024; Pane et al., 2024).

Algoritma ini dipilih karena memiliki kompleksitas perhitungan yang relatif rendah, tidak dipengaruhi oleh panjang dokumen, serta banyak digunakan dalam penelitian text mining dan information retrieval (Abdurrahman et al., 2024; Nasrullah, 2024). Dalam penelitian ini, Cosine Similarity difungsikan sebagai algoritma inti dalam menentukan perolehan skor jawaban esai yang diberikan oleh peserta ujian.

6) Representasi Teks Menggunakan TF-IDF

Term Frequency–Inverse Document Frequency (TF-IDF) merupakan metode pembobotan kata yang bertugas mengubah data teks ke dalam bentuk representasi numerik. Metode ini memberikan nilai bobot yang lebih tinggi kepada kata-kata yang memiliki tingkat kemunculan yang sering dalam suatu dokumen tertentu, namun relatif jarang

dijumpai pada dokumen-dokumen lainnya (Kerina Putri & Laili Cahyani, 2026; Pradana et al., 2024).

Dalam penelitian ini, TF-IDF diterapkan untuk membangun representasi vektor kata dari jawaban siswa maupun kunci jawaban sebagai tahapan awal sebelum proses penghitungan Cosine Similarity dilakukan. Representasi vektor yang baik akan meningkatkan akurasi pengukuran tingkat kemiripan antar kata.

7) Kecerdasan Buatan dalam Penilaian Text

Kecerdasan buatan (Artificial Intelligence) memungkinkan suatu sistem untuk melakukan analisis data secara mandiri dan menghasilkan keputusan berdasarkan pola-pola yang berhasil diidentifikasi dari data yang diproses. Dalam bidang pendidikan, AI telah banyak digunakan untuk mendukung proses evaluasi pembelajaran, termasuk penilaian esai otomatis (Maulidya Prastita Syah et al., 2025; Pradani & Suadaa, 2023).

Penelitian ini menggunakan pendekatan AI untuk memperkuat akurasi algoritma text similarity sehingga sistem CBT yang dikembangkan dapat melakukan penilaian esai dengan lebih cerdas dan efisien.

8) Penelitian Terdahulu

Penelitian yang dilaksanakan oleh (Pradani & Suadaa, 2023) merancang dan mengembangkan sistem Automated Essay Scoring dengan memanfaatkan pendekatan Semantic Textual Similarity berbasis Transformer. Hasil yang diperoleh menunjukkan bahwa sistem tersebut mampu menghasilkan penilaian yang memiliki kedekatan tinggi dengan hasil evaluasi yang dilakukan oleh manusia.

Penelitian yang dilaksanakan oleh (Maulidya Prastita Syah et al., 2025) mengkaji perbandingan representasi teks dengan memanfaatkan TF-IDF dan BERT dalam proses penilaian jawaban yang berbasis teks secara otomatis dengan menggunakan pendekatan Cosine Similarity. Temuan penelitian tersebut mengungkapkan bahwa representasi berbasis BERT terbukti lebih unggul dalam menangkap pemahaman konteks dibandingkan TF-IDF, meskipun demikian TF-IDF masih memiliki kelebihan tersendiri dari segi efisiensi komputasi dan kemudahan dalam proses implementasinya. Oleh karena itu temuan tersebut menunjukkan bahwa representasi teks memiliki pengaruh yang signifikan terhadap kualitas penilaian otomatis.

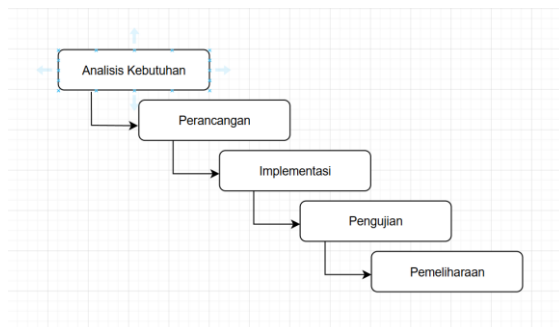
Penelitian (Rokhman et al., 2025) mengimplementasikan NLP dan Cosine Similarity pada sistem penilaian esai secara otomatis. Temuan penelitian memperlihatkan tingkat kemiripan rata-rata sebesar 88,2% dan nilai MAE sebesar 2 sehingga mampu menghasilkan penilaian yang mendekati koreksi manual.

2. METODE

2.1. Desain Penelitian

Penelitian ini menggunakan pendekatan pengembangan sistem yang mengacu pada metode Waterfall. Pemilihan metode ini didasarkan pada

karakteristiknya yang memiliki alur tahapan terstruktur dan sistematis, sehingga dinilai sesuai untuk pengembangan perangkat lunak yang kebutuhan sistemnya sudah dapat didefinisikan secara jelas sejak awal proses penelitian berlangsung (Al Azfar & Anggita, 2024). Tahapan dalam penelitian ini meliputi analisis kebutuhan, desain sistem, implementasi, pengujian, serta pemeliharaan. Sistem yang dikembangkan berupa aplikasi Computer Based Test (CBT) yang berbasis web yang memiliki kemampuan untuk melakukan penilaian secara otomatis terhadap soal esai dengan memanfaatkan algoritma Text Similarity berbasis kecerdasan buatan Artificial Intelligence (AI).



Gambar 1. Tahapan metode waterfall

2.2. Pengumpulan Data

Data yang digunakan dalam penelitian ini terdiri atas data pengguna, data soal esai, data kunci jawaban, serta data jawaban yang diberikan oleh peserta ujian. Pengumpulan data dilaksanakan melalui dua metode, yaitu pengamatan langsung terhadap proses evaluasi pembelajaran yang masih bergantung pada penilaian manual, serta penelusuran literatur dari berbagai sumber penelitian yang berhubungan dengan Automated Essay Scoring, Natural Language Processing (NLP), TF-IDF, dan Cosine Similarity. Data jawaban siswa dan kunci jawaban selanjutnya dijadikan sebagai data utama dalam proses penghitungan tingkat kemiripan teks guna menghasilkan nilai secara otomatis.

2.3. Arsitektur dan Prosedur Sistem

Pengembangan sistem dimulai dengan tahap analisis kebutuhan fungsional dan nonfungsional yang melibatkan dua kelompok pengguna, yakni guru dan siswa. Guru memiliki hak akses untuk mengelola soal esai, kunci jawaban, data peserta, serta melihat hasil ujian. Sementara itu, siswa dapat mengikuti ujian dan melihat hasil evaluasi yang diperoleh.

Tahap perancangan sistem meliputi pembangunan arsitektur aplikasi, perancangan struktur basis data, desain tampilan antarmuka pengguna, serta penyusunan mekanisme penilaian esai yang dilakukan secara otomatis. Pada tahap implementasi, jawaban siswa beserta kunci jawaban diproses menggunakan teknik Natural Language Processing (NLP) yang melibatkan sejumlah tahapan preprocessing, di antaranya case folding, tokenizing, stopword removal, dan stemming (Sihombing, 2022).

Setelah proses preprocessing rampung dilaksanakan, teks selanjutnya dikonversi ke dalam bentuk numerik dengan menerapkan metode Term Frequency-Inverse Document Frequency (TF-IDF) untuk menetapkan bobot pada masing-masing kata berdasarkan tingkat signifikansinya dalam dokumen tersebut (Kerina Putri & Laili Cahyani, 2026; Nasrullah, 2024).

2.4. Metrik Evaluasi

Evaluasi sistem dilaksanakan dengan menggunakan metode Black Box Testing yang memusatkan perhatian pada pengujian fungsionalitas sistem secara menyeluruh tanpa mengacu pada struktur kode program yang terdapat di dalamnya (Jannah, 2023). Pengujian dilaksanakan pada semua fitur pokok yang ada dalam sistem, meliputi proses login, manajemen soal, pelaksanaan ujian, penilaian secara otomatis, serta penampilan hasil evaluasi. Sistem dinyatakan berhasil apabila semua fungsi menghasilkan keluaran yang sesuai dengan kebutuhan pengguna serta spesifikasi sistem yang telah dirancang dan ditetapkan sebelumnya. Di samping itu, hasil penghitungan Cosine Similarity dijadikan sebagai acuan dalam menentukan nilai otomatis yang diberikan kepada setiap peserta ujian.

3. PEMBAHASAN

3.1. Analisis Kebutuhan Sistem

Tahap analisis kebutuhan dijalankan dengan tujuan untuk mengetahui dan mendefinisikan kebutuhan pengguna terhadap sistem Computer Based Test (CBT) yang dilengkapi dengan kemampuan penilaian otomatis pada soal esai menggunakan algoritma Text Similarity berbasis kecerdasan buatan (AI). Berdasarkan temuan dari observasi dan penelusuran literatur yang telah dilaksanakan, proses penilaian soal esai yang hingga kini masih dijalankan secara manual terbukti memerlukan waktu yang cukup panjang, terlebih ketika jumlah peserta ujian dalam jumlah besar. Selain itu, proses koreksi secara manual juga berpotensi memunculkan unsur subjektivitas dalam pemberian nilai, mengingat penilaian sangat bergantung pada sudut pandang dan persepsi masing-masing pengajar (Rokhman et al., 2025).

Sistem yang dikembangkan dalam penelitian ini dirancang agar dapat dioperasikan oleh dua pengguna, yakni guru dan siswa. Guru berperan dalam mengelola soal esai, kunci jawaban, serta melihat hasil ujian peserta. Sementara itu, siswa berperan sebagai peserta ujian yang dapat mengerjakan soal dan melihat hasil penilaian yang diperoleh setelah ujian selesai dilaksanakan.

Kebutuhan fungsional sistem meliputi pengelolaan soal esai, pengelolaan kunci jawaban, pelaksanaan ujian berbasis komputer, penilaian otomatis jawaban esai, dan penyajian laporan hasil ujian. Sedangkan kebutuhan nonfungsional sistem mencakup aspek kemudahan penggunaan, keamanan data, kecepatan

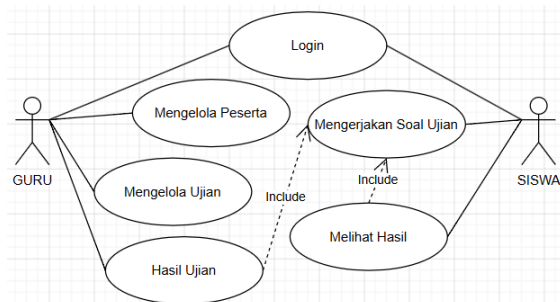
dalam proses penilaian, serta kapabilitas sistem untuk diakses melalui web (Ulum & Hidayat, 2024).

3.2. Perancangan Sistem

Perancangan sistem dijalankan berdasarkan kebutuhan yang telah diidentifikasi pada tahap analisis yang dilakukan sebelumnya. Proses perancangan ini meliputi pemodelan sistem dengan memanfaatkan Use Case Diagram serta perancangan struktur basis data menggunakan Entity Relationship Diagram (ERD).

1) Use Case Diagram

Use Case Diagram menggambarkan pola hubungan dan interaksi yang berlangsung antara pengguna dengan sistem. Seperti ditunjukkan pada Gambar 2, terdapat dua aktor utama yang berperan dalam sistem ini, yakni guru dan siswa. Guru dapat mengelola peserta, mengelola ujian, melihat hasil ujian, dan mencetak laporan nilai. Siswa dapat mengikuti ujian dan melihat hasil penilaian yang diperoleh.



Gambar 2. Use case diagram

2) Entity Relationship Diagram (ERD)

ERD diterapkan untuk memetakan hubungan antar data yang ada di dalam sistem. Gambar 3 menunjukkan entitas pokok yang digunakan dalam perancangan ini meliputi tabel guru, siswa, soal, jawaban, ujian, dan hasil penilaian. Hubungan antar entitas dirancang agar proses penyimpanan data ujian beserta hasil penilaian dapat dijalankan secara terpadu dan terintegrasi.



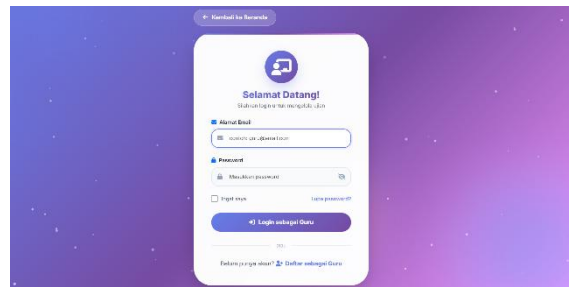
Gambar 3. Entity relationship diagram

3.3. Implementasi Sistem

Tahap implementasi menghasilkan sebuah aplikasi CBT berbasis web yang dapat digunakan oleh guru maupun siswa dalam melaksanakan kegiatan evaluasi hasil pembelajaran secara digital. Sistem ini dikembangkan dengan pendekatan berbasis web sehingga memungkinkan untuk diakses melalui berbagai perangkat yang terhubung dengan jaringan internet.

1) Halaman Login Guru

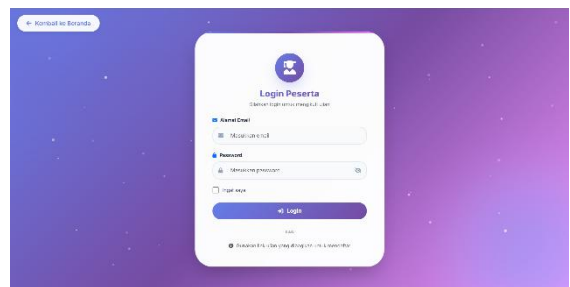
Gambar 4 memperlihatkan halaman login guru yang berfungsi sebagai proses verifikasi identitas yang wajib dilewati oleh guru sebelum dapat masuk dan mengakses sistem. Guru diwajibkan menginputkan alamat email dan kata sandi yang benar dan terdaftar pada kolom yang tersedia, selanjutnya mengklik tombol login untuk dapat mengakses dan masuk ke dalam sistem.



Gambar 4. Tampilan login guru

2) Halaman Login Siswa

Halaman login siswa berfungsi sebagai proses verifikasi identitas yang wajib dijalani oleh siswa sebelum dapat membuka dan mengakses sistem. Seperti yang terlihat pada Gambar 5 Siswa diwajibkan menginputkan alamat email serta kata sandi yang tepat dan telah terdaftar pada kolom yang tersedia, lalu menekan tombol login untuk dapat mengakses dan masuk ke dalam sistem.

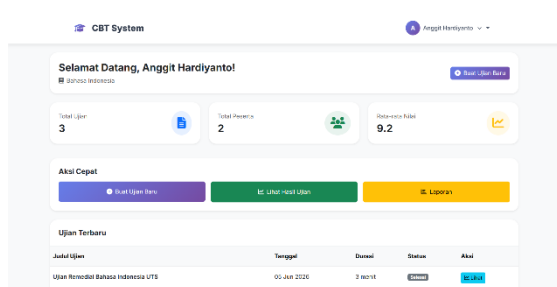


Gambar 5. Tampilan login siswa

3) Halaman Dashboard Guru

Dashboard guru menampilkan informasi mengenai jumlah ujian, jumlah peserta ujian, serta hasil ujian yang telah dilaksanakan. Pada Gambar 6 guru juga

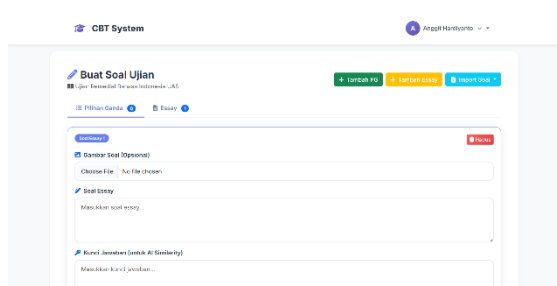
dapat mengakses menu pengelolaan soal dan hasil penilaian.



Gambar 6. Tampilan dashboard guru

4) Halaman Buat Soal

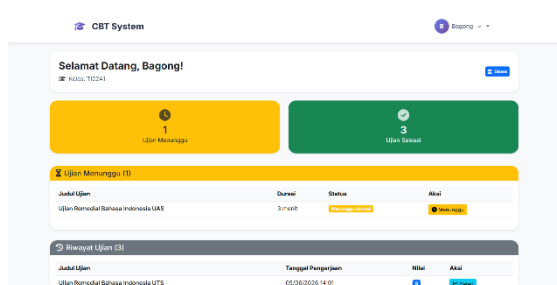
Pada Gambar 7 merupakan tempat untuk guru membuat soal untuk peserta ujian/siswa. Di halaman ini disediakan input soal, gambar dan kunci jawaban serta nilai jawaban. Pada halaman ini turut tersedia tombol untuk mengimpor soal dalam format file excel.



Gambar 7. Tampilan buat soal

5) Halaman Dashboard Siswa

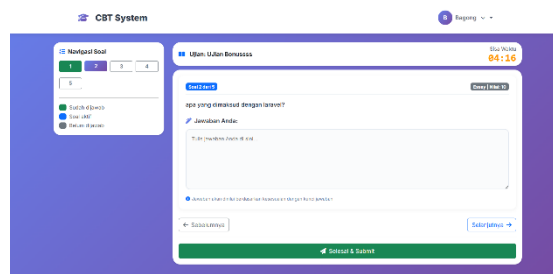
Gambar 8 memperlihatkan informasi terkait jumlah ujian yang akan segera dilaksanakan serta jumlah ujian yang sudah berhasil diselesaikan oleh siswa. Pada halaman ini juga terdapat tabel yang menampilkan daftar ujian yang sedang menunggu ujian yang terdapat tombol untuk mulai mengerjakan serta terdapat juga tabel riwayat ujian.



Gambar 8. Tampilan dashboard siswa

6) Halaman Ujian Siswa

Halaman ujian digunakan oleh siswa untuk mengerjakan soal esai yang telah disediakan. Jawaban yang dikirimkan akan tersimpan pada basis data dan diproses secara otomatis oleh sistem. Pada Gambar 9 ini juga disediakan navigasi soal dan waktu untuk mengerjakan ujian, dan jika waktu habis maka ujian otomatis berhenti dan diarahkan untuk melihat hasil ujian.



Gambar 9. Tampilan ujian siswa

7) Implementasi Algoritma Penilaian Esai

Penilaian otomatis dilakukan menggunakan tahapan NLP, TF-IDF, dan Cosine Similarity. Contoh implementasi ditunjukkan pada data berikut.

Tabel 1. Contoh kunci jawaban dan jawaban siswa

Kunci Jawaban	Jawaban Siswa
Kecerdasan buatan adalah teknologi yang memungkinkan komputer meniru kemampuan berpikir manusia.	Kecerdasan buatan merupakan teknologi yang membuat komputer dapat meniru cara berpikir manusia.

Tabel 2. Hasil preprocessing

Tahapan	Kunci Jawaban	Jawaban siswa
Case Folding	kecerdasan buatan adalah teknologi yang memungkinkan komputer meniru kemampuan berpikir manusia	kecerdasan buatan merupakan teknologi yang membuat komputer dapat meniru cara berpikir manusia
Tokenizing	[kecerdasan, buatan, adalah, teknologi, yang, memungkinkan, komputer, meniru, kemampuan, berpikir, manusia]	[kecerdasan, buatan, merupakan, teknologi, yang, membuat, komputer, dapat, meniru, cara, berpikir, manusia]
Stopword Removal	[kecerdasan, buatan, teknologi, memungkinkan, komputer, meniru,	[kecerdasan, teknologi, membuat, komputer, meniru,

	kemampuan, berpikir, manusia]	berpikir, manusia]
Stemming	[cerdas, buat, teknologi, mungkin, komputer, tiru, mampu, pikir, manusia]	[cerdas, buat, teknologi, buat, komputer, tiru, pikir, manusia]

Setelah proses preprocessing, dilakukan pembobotan menggunakan metode Term Frequency-Inverse Document Frequency (TF-IDF) untuk menghasilkan representasi numerik dari setiap dokumen sehingga kata-kata yang dianggap penting memiliki bobot yang lebih besar (Kerina Putri & Laili Cahyani, 2026; Pradana et al., 2024). Selanjutnya, vektor TF-IDF dari jawaban siswa dan kunci jawaban dibandingkan menggunakan algoritma Cosine Similarity untuk mengukur tingkat kemiripan kedua dokumen. Perhitungan kemiripan dilakukan menggunakan Persamaan (3), dengan nilai berada pada rentang 0 hingga 1, di mana nilai yang mendekati 1 menunjukkan tingkat kesesuaian jawaban yang semakin tinggi (Abdurrahman et al., 2024; Rokhman et al., 2025).

Pembobotan kata menggunakan metode TF-IDF dihitung menggunakan persamaan (1).

$$TF - IDF(t, d) = TF(t, d) \times IDF(t) \quad (1)$$

Nilai IDF dihitung menggunakan Persamaan (2).

$$IDF(t) = \log\left(\frac{N}{DF(t)}\right) \quad (2)$$

Setelah diperoleh representasi vektor menggunakan TF-IDF, tingkat kemiripan antara jawaban siswa dan kunci jawaban dihitung menggunakan algoritma Cosine Similarity sebagaimana ditunjukkan pada Persamaan (3).

$$\text{Cosine Similarity}(A, B) = \frac{A \cdot B}{|A| \times |B|} \quad (3)$$

Berdasarkan hasil pembobotan TF-IDF yang dihasilkan oleh sistem, diperoleh nilai Cosine Similarity Cosine antara vektor jawaban siswa dan kunci jawaban. Hasil perhitungan menghasilkan nilai kemiripan sebesar 0,7379 atau 73,79%, sehingga sistem memberikan nilai akhir sebesar 74% kepada siswa. Hasil tersebut mengindikasikan bahwa jawaban siswa memiliki tingkat kecocokan yang tinggi terhadap kunci jawaban (Abdurrahman et al., 2024; Pane et al., 2024).

8) Halaman Hasil Ujian

Gambar 10 memperlihatkan nilai yang diperoleh siswa setelah mengikuti ujian serta terdapat table peringkat antar siswa setelah ujian selesai.

Peringkat	Nama	Kelas	Nilai	Salah
1	Eugenia	TE2241	32	21/25/18
2	Eugeng	TE2241	32	21/25/20

Gambar 10. Tampilan hasil ujian

3.4. Pengujian Sistem

Pengujian terhadap sistem dilakukan menggunakan metode Black Box Testing dengan tujuan memastikan bahwa setiap fungsi yang ada di dalamnya berjalan sesuai dengan kebutuhan yang sudah ditentukan sebelumnya (Jannah, 2023).

Tabel 3. Pengujian fungsionalitas sistem

Skenario Pengujian	Output yang Diharapkan	Hasil Pengujian	Status
Login pengguna (data valid)	Sistem menampilkan tampilan dashboard yang disesuaikan dengan peran masing-masing pengguna.	Sistem menampilkan tampilan dashboard yang sesuai dengan hak akses yang dimiliki pengguna.	Valid
Login pengguna (data tidak valid)	Sistem menampilkan notifikasi pesan kesalahan.	Sistem menampilkan pemberitahuan bahwa proses login tidak berhasil dilakukan.	Valid
Logout pengguna	Sistem kembali menuju ke halaman login	Sistem berhasil melakukan proses keluar (logout) dari akun	Valid
Membuat soal	Data soal seperti gambar, soal, kunci jawaban dan nilai jawaban berhasil disimpan ke dalam database	Data soal berhasil disimpan ke dalam database.	Valid
Membuat soal (masih terdapat data yang belum diisi secara lengkap)	Sistem menampilkan sebuah pesan peringatan, dan data soal tidak bisa tersimpan ke dalam database	Sistem menampilkan pesan pemberitahuan dan data soal gagal disimpan ke dalam database.	Valid
Mengerjakan ujian	Sistem menampilkan soal yang tersimpan di database kemudian sistem berhasil menyimpan jawaban	Sistem menampilkan soal dan mampu menyimpan jawaban ke database	Valid

Menampilkan hasil nilai	Sistem menghitung dan menampilkan jumlah nilai masing-masing peserta/siswa	Sistem menghitung dan menampilkan jumlah nilai	Valid
-------------------------	--	--	-------

Mengacu pada hasil pengujian fungsional dengan metode Black Box yang ditampilkan pada Tabel 4, dapat diketahui bahwa seluruh fitur utama sistem Computer Based Test telah berfungsi sesuai dengan kebutuhan yang telah ditentukan sebelumnya. Pada setiap skenario pengujian, terlihat adanya kesesuaian antara input yang diberikan dengan output yang dihasilkan oleh sistem. Hasil pengujian juga menunjukkan bahwa sistem mampu menangani kesalahan input dengan baik melalui mekanisme validasi dan notifikasi, sehingga dapat meminimalkan kesalahan penggunaan oleh pengguna.

Selain pengujian fungsional, dilakukan evaluasi terhadap kinerja algoritma penilaian esai menggunakan metode Mean Absolute Error (MAE). Pengujian dilakukan dengan membandingkan nilai yang diberikan guru dengan nilai yang dihasilkan sistem terhadap sepuluh jawaban esai siswa.

Tabel 4. Perbandingan nilai guru dan sistem

No Absen Siswa	Nilai Guru	Nilai Sistem	Selisih
1	85	83	91%
2	78	75	87%
3	90	88	93%
4	70	68	86%
5	62	61	74%
6	88	86	90%
7	76	74	88%
8	82	80	89%
9	79	77	87%
10	84	82	91%

Perhitungan MAE:

$$\begin{aligned}
 MAE &= \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{n} \\
 &= \frac{2 + 3 + 2 + 2 + 1 + 2 + 2 + 2 + 2 + 2}{10} \quad (4) \\
 &= \frac{20}{10} = 2
 \end{aligned}$$

Berdasarkan pengujian terhadap sepuluh jawaban esai yang salah satu data uji menghasilkan nilai kemiripan sebesar 74% yang menunjukkan bahwa terdapat perbedaan penggunaan kata atau informasi yang menyebabkan tingkat kesesuaian jawaban lebih rendah dibandingkan data lainnya. Sementara itu,

hasil evaluasi menggunakan metode Mean Absolute Error (MAE) menunjukkan nilai sebesar 2, yang berarti rata-rata selisih antara penilaian sistem dan penilaian guru hanya sebesar dua poin. Hasil tersebut menunjukkan bahwa sistem memiliki kemampuan yang cukup baik dalam melakukan penilaian otomatis terhadap jawaban esai dan mampu menghasilkan nilai yang mendekati penilaian manual guru.

Dengan demikian, hasil pengujian Black Box dan perhitungan MAE menunjukkan bahwa sistem telah memenuhi seluruh aspek fungsionalitas yang dipersyaratkan dan layak untuk diterapkan sebagai sarana pendukung proses evaluasi di berbagai lingkungan pendidikan, baik di tingkat sekolah, perguruan tinggi, maupun di lingkungan perusahaan.

3.5. Hasil

Hasil penelitian menunjukkan bahwa algoritma Text Similarity berbasis AI dapat diterapkan pada sistem CBT untuk melakukan penilaian esai secara otomatis. Serangkaian tahapan NLP yang meliputi case folding, tokenizing, stopword removal, dan stemming terbukti dapat meningkatkan kualitas representasi teks secara signifikan, sehingga proses penghitungan tingkat kemiripan antar teks dapat berjalan dengan lebih optimal (Rokhman et al., 2025; Sihombing, 2022).

Penerapan TF-IDF dapat memberikan bobot yang lebih signifikan pada kata-kata yang dinilai memiliki tingkat kepentingan tinggi dalam suatu dokumen, sehingga dapat meningkatkan kualitas representasi teks sebelum proses penghitungan kemiripan dilaksanakan. (Kerina Putri & Laili Cahyani, 2026; Nasrullah, 2024). Lebih lanjut, algoritma Cosine Similarity terbukti dapat menghitung tingkat kesesuaian antara jawaban siswa dengan kunci jawaban secara efektif, tanpa terpengaruh oleh panjang atau pendeknya dokumen yang dibandingkan (Abdurrahman et al., 2024; Pane et al., 2024).

Temuan dalam penelitian ini sesuai dengan hasil penelitian (Pradani & Suadaa, 2023) yang menunjukkan bahwa pendekatan Automated Essay Scoring dapat menghasilkan penilaian yang mendekati evaluasi manusia. Penelitian ini juga mendukung hasil penelitian (Abdurrahman et al., 2024; Rokhman et al., 2025) yang mengungkapkan bahwa perpaduan antara NLP, TF-IDF, dan Cosine Similarity dapat meningkatkan tingkat objektivitas serta efisiensi dalam proses penilaian esai. Perbedaan penelitian ini terletak pada integrasi proses penilaian otomatis ke dalam sistem CBT berbasis web sehingga proses ujian, penilaian, dan penyajian hasil dapat dilakukan dalam satu platform yang terintegrasi. Dengan demikian, sistem yang dikembangkan dapat membantu guru mengurangi waktu koreksi, meningkatkan konsistensi penilaian,

serta mendukung transformasi digital dalam proses evaluasi pembelajaran.

4. KESIMPULAN

Mengacu pada hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa sistem Computer Based Test (CBT) dengan kemampuan penilaian otomatis soal esai menggunakan algoritma Text Similarity berbasis kecerdasan buatan (Artificial Intelligence/AI) berhasil dikembangkan dengan menerapkan metode Waterfall, serta mampu mengintegrasikan proses pelaksanaan ujian, pengelolaan soal esai, dan penilaian otomatis secara terpadu dalam satu platform berbasis web. Proses penilaian dijalankan melalui serangkaian tahapan Natural Language Processing (NLP) yang meliputi case folding, tokenizing, stopword removal, dan stemming, yang selanjutnya diikuti dengan proses pembobotan menggunakan TF-IDF serta perhitungan tingkat kemiripan teks dengan memanfaatkan algoritma Cosine Similarity. Hasil pengujian menggunakan metode Black Box Testing menunjukkan bahwa semua fitur yang tersedia dalam sistem telah berjalan sesuai dengan kebutuhan yang telah ditetapkan. Sementara itu, hasil penilaian yang dihasilkan oleh sistem menunjukkan tingkat kedekatan yang baik dengan penilaian yang diberikan oleh guru, sehingga sistem ini dapat menghasilkan penilaian secara lebih cepat, konsisten, dan objektif dibandingkan dengan proses koreksi yang dilakukan secara manual. Temuan penelitian ini memperkuat pemanfaatan NLP, TF-IDF, dan Cosine Similarity dalam pengembangan Automated Essay Scoring, sekaligus memberikan kontribusi praktis dalam mendukung digitalisasi evaluasi pembelajaran melalui penerapan teknologi AI pada sistem CBT. Untuk penelitian selanjutnya, disarankan melakukan pengembangan dengan memanfaatkan model bahasa yang lebih canggih seperti BERT atau IndoBERT serta melakukan pengujian pada jumlah data dan variasi mata pelajaran yang lebih luas guna meningkatkan akurasi dan kemampuan generalisasi sistem.

PUSTAKA

- Abdurrahman, A., Sucipto, H., Permadi, G. S., & Vitadiar, T. Z. (2024). Rancang bangun aplikasi penilaian esai otomatis menggunakan algoritma Cosine Similarity. *Inovate*, 9(1).
- Al Azfar, N. A., & Anggita, S. D. (2024). Penerapan metode Waterfall pada sistem informasi E-Rapor. *Information System Journal (INFOS)*, 7(1), 45–55.
- Jannah, N. (2023). Analisis pengembangan sistem informasi batik Madura menggunakan metode Waterfall dan Black Box Testing. *Jurnal Satya Informatika*, 8(1), 64–81.
- Nasrullah, A. H. (2024). Integrasi TF-IDF dan algoritma Cosine Similarity untuk deteksi tingkat kemiripan judul penelitian (studi kasus mahasiswa Fakultas Ilmu Komputer UNISAN Gorontalo). *Information Technology Education Journal*, 3(3)..
- Pane, E. S., Hardianto, R., & Choriah, W. (2024). Sistem penilaian ujian esai secara otomatis dengan algoritma text mining Cosine Similarity penunjang pembelajaran. *ZONasi Jurnal Sistem Informasi*, 6(2).
- Pradana, M. G., Irzavika, N., & Maulana, N. (2024). Deteksi kemiripan dokumen menggunakan Cosine Similarity berdasarkan representasi teks Count Vectorizer dan TF-IDF. *Indonesian Journal of Business Intelligence (IJUBI)*, 7(2).
- 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).
- Putri, K., Nanda, & Laili. (2025). Penerapan algoritma TF-IDF dan Cosine Similarity untuk query pencarian soal mata pelajaran sosiologi SMA. *Jurnal Teknologi Informasi: Jurnal Keilmuan dan Aplikasi Bidang Teknik Informatika*, 20(1).
- Rokhman, N., Maulan, P. A., & Wirahuda, N. A. (2025). Analisis penilaian esai secara otomatis menggunakan Natural Language Processing (NLP) dan Cosine Similarity. *Go Infotech*, 31(1).
- Sihombing, D. O. (2022). Implementasi Natural Language Processing dan algoritma Cosine Similarity dalam penilaian ujian esai otomatis. *Jurnal Sistem Komputer dan Informatika (JSON)*, 4(2).
- Syah, M. P., Wardani, A. P., Idhom, M., & Trimono. (2025). Perbandingan representasi teks TF-IDF dan BERT terhadap akurasi Cosine Similarity dalam penilaian otomatis jawaban berbasis teks. *Data Sciences Indonesia (DSI)*, 5(1).
- Ulum, M. N., & Hidayat, S. (2024). Rancang bangun Computer Based Test (CBT) berbasis web. *Jurnal Indonesia: Manajemen Informatika dan Komunikasi*, 5(2).