

Implementasi *Term Frequency* dan *Cosine Similarity* pada Aplikasi Koreksi Otomatis Jawaban Esai Berbasis Web

Joko Suwarno ¹, Galuh Saputri ^{2*}, Willis Puspita Sari ³, Didik Kusuma Rahmat ⁴

^{1*,2,3,4} Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Pamulang, Kota Tangerang Selatan, Provinsi Banten, Indonesia.

Email: dosen02522@unpam.ac.id ^{1*}, dosen02693@unpam.ac.id ², dosen02527@unpam.ac.id ³, didik.kusuma3@gmail.com ⁴

Histori Artikel:

Dikirim 9 September 2026; *Diterima dalam bentuk revisi* 14 September 2026; *Diterima* 20 September 2026; *Diterbitkan* 30 September 2026. Semua hak dilindungi oleh Lembaga Penelitian dan Pengabdian Masyarakat (LPPM) STM IK Indonesia Banda Aceh.

Abstrak

Penilaian jawaban esai secara manual membutuhkan waktu dan tenaga karena guru harus memeriksa jawaban peserta didik satu per satu berdasarkan kunci jawaban. Penelitian ini bertujuan mengimplementasikan Term Frequency (TF) dan Cosine Similarity pada aplikasi koreksi otomatis jawaban esai berbasis web. Pengembangan sistem menggunakan Software Development Life Cycle (SDLC) model Waterfall, sedangkan pengolahan teks dilakukan melalui tahap case folding, tokenization, filtering, dan stemming. Hasil preprocessing direpresentasikan dalam bentuk frekuensi term dan selanjutnya dibandingkan menggunakan Cosine Similarity. Demonstrasi perhitungan menggunakan satu pasangan kunci jawaban dan jawaban siswa menghasilkan nilai Cosine Similarity sebesar 0,878 (87,8%), yang menunjukkan tingkat kemiripan tekstual antara kedua dokumen; nilai ini bukan akurasi sistem karena pengujian algoritma pada penelitian ini masih terbatas pada satu pasangan dokumen dan belum dibandingkan dengan penilaian guru. Pengujian fungsional menggunakan Black Box Testing menunjukkan bahwa fungsi utama aplikasi dapat berjalan sesuai skenario yang dirancang. Hasil penelitian menunjukkan bahwa TF dan Cosine Similarity dapat diimplementasikan sebagai mekanisme pengukuran kemiripan tekstual dalam aplikasi koreksi esai berbasis web. Validasi menggunakan kumpulan jawaban siswa yang lebih besar dan perbandingan dengan penilaian guru diperlukan untuk mengevaluasi kinerja metode secara lebih komprehensif.

Kata Kunci: Term Frequency; Cosine Similarity; Koreksi Esai; Waterfall; Web.

Abstract

Manually grading essay answers is time-consuming and labor-intensive, as teachers must review each student's response individually against an answer key. This study aims to implement Term Frequency (TF) and Cosine Similarity within a web-based automated essay grading application. The system was developed using the Waterfall Software Development Life Cycle (SDLC) model, with text processing involving case folding, tokenization, filtering, and stemming stages. Preprocessing results were represented as term frequencies and subsequently compared using Cosine Similarity. A sample calculation using a single pair—an answer key and a student response—yielded a Cosine Similarity score of 0.878 (87.8%), indicating the level of textual similarity between the two documents; this figure does not represent the system's overall accuracy, as algorithm testing in this study was limited to a single document pair and did not include a comparison with teacher assessments. Functional testing using Black Box Testing demonstrated that the application's core functions operated according to the designed scenarios. The results indicate that TF and Cosine Similarity can be implemented as a mechanism for measuring textual similarity in web-based essay grading applications. Validation using a larger dataset of student answers and comparisons with teacher assessments is required to evaluate the method's performance more comprehensively.

Keyword: Term Frequency; Cosine Similarity; Essay Grading; Waterfall; Web.

1. Pendahuluan

Evaluasi pembelajaran merupakan salah satu komponen penting dalam mengukur tingkat pemahaman dan pencapaian kompetensi peserta didik. Salah satu bentuk evaluasi yang dapat digunakan adalah soal esai yang memungkinkan peserta didik menyampaikan pemahaman, argumentasi, serta penalaran dalam bentuk uraian. Namun, proses koreksi jawaban esai secara manual membutuhkan waktu dan tenaga karena guru harus membaca serta mencocokkan jawaban peserta didik dengan kunci jawaban satu per satu. Permasalahan tersebut menjadi semakin kompleks ketika jumlah peserta didik dan soal yang harus diperiksa meningkat. Selain melaksanakan proses pembelajaran, guru juga memiliki berbagai tugas administratif sehingga proses koreksi manual dapat menjadi beban tambahan. Kondisi tersebut menunjukkan perlunya pemanfaatan teknologi untuk meningkatkan efisiensi dan konsistensi proses penilaian.

Perkembangan *Natural Language Processing* (NLP) memberikan peluang untuk mengembangkan *Automated Essay Scoring* (AES), yaitu sistem yang mampu melakukan penilaian jawaban berbasis teks secara otomatis (Andhini *et al.*, 2025; Rokhman *et al.*, 2025). Penelitian Lahitani (2022) menunjukkan bahwa kombinasi *Term Frequency-Inverse Document Frequency* dan *Cosine Similarity* dapat digunakan untuk melakukan penilaian esai multisoal dengan membandingkan representasi jawaban peserta didik terhadap jawaban acuan. Penelitian Kurniadi *et al.* (2023) mengembangkan sistem *Automated Short Answer Grading* dengan *casefolding*, *tokenizing*, *stopword removal*, dan *Cosine Similarity*; hasil penelitian tersebut memperoleh rata-rata kemiripan sebesar 85,4% serta korelasi yang tinggi antara hasil sistem dan penilaian instruktur, dengan nilai Kendall's W masing-masing 0,885 dan 0,883.

Penelitian Dillah *et al.* (2026) menunjukkan bahwa *fine-tuning* SBERT menggunakan *triplet loss* mampu meningkatkan representasi semantik teks bahasa Indonesia. Pengujian pada 3.000 pasangan kalimat sinonim dan nonsinonim menghasilkan akurasi 92% dan F1-score 0,92, lebih tinggi dibandingkan IndoBERT, SBERT *pralatih*, FastText, dan TF-IDF. Hasil ini menunjukkan bahwa pendekatan berbasis *embedding* kontekstual lebih efektif dalam menangkap variasi makna dan konteks bahasa Indonesia dibanding metode berbasis leksikal. Penerapan kombinasi TF-IDF dan *Cosine Similarity* oleh Saputra *et al.* (2024) untuk menentukan rumpun jabatan menunjukkan bahwa metode ini dapat digunakan untuk mengukur kedekatan antarrepresentasi teks; temuan ini memperkuat relevansi TF-IDF dan *Cosine Similarity* sebagai pendekatan yang relatif sederhana dan dapat diimplementasikan pada sistem pengolahan teks.

Pendekatan berbasis TF dan *Cosine Similarity* memiliki keterbatasan karena kemiripan yang dihasilkan sangat dipengaruhi oleh kesamaan kata atau fitur leksikal. Jawaban yang memiliki makna sama tetapi menggunakan struktur kalimat dan kosakata berbeda berpotensi memperoleh nilai kemiripan yang lebih rendah (Anggraeni *et al.*, 2026; Suryani & Edy, 2020). Keterbatasan tersebut mendorong penggunaan pendekatan berbasis representasi semantik. Chamidah *et al.* (2022) menggunakan *Sentence-BERT* (SBERT) untuk memperoleh representasi semantik jawaban esai dan memperoleh rata-rata MAE sebesar 0,26 serta korelasi Pearson sebesar 0,78 terhadap penilaian evaluator manusia. Sejalan dengan itu, Gudya *et al.* (2022) mengkaji pengaruh *text summarization* dalam penilaian esai dan menunjukkan bahwa pengolahan karakteristik teks merupakan bagian penting dalam pengembangan AES. Sitorus *et al.* (2026) kemudian mengembangkan pendekatan *hybrid* yang menggabungkan *Sentence-BERT*, Jaccard, Levenshtein, *keyword coverage*, dan *length penalty* untuk penilaian esai siswa SMA, dengan hasil MAE 3,58 dan RMSE 4,48.

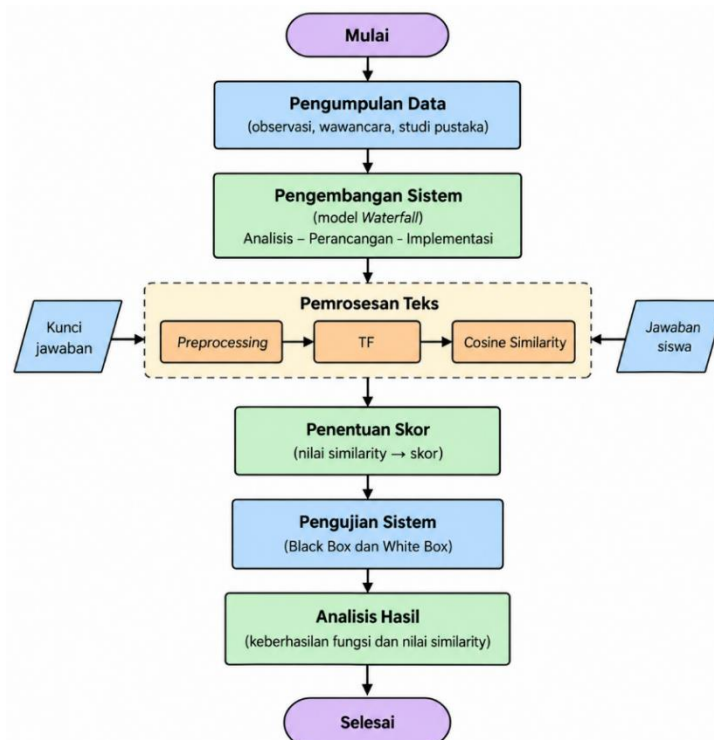
Perkembangan tersebut menunjukkan adanya *research gap*. Penelitian terdahulu cenderung berfokus pada pengembangan dan perbandingan metode untuk meningkatkan kesesuaian skor otomatis dengan penilaian manusia, sedangkan kebutuhan praktis di lingkungan sekolah tidak hanya berkaitan dengan algoritma penilaian, tetapi juga bagaimana metode tersebut diintegrasikan ke dalam proses ujian secara menyeluruh. Penelitian Permata *et al.* (2025) mengembangkan sistem evaluasi esai berbasis NLP menggunakan TF-IDF, *Cosine Similarity*, dan Manhattan Distance, tetapi penelitian tersebut berfokus pada evaluasi metode dan metrik kemiripan. Berdasarkan hal tersebut, terdapat kebutuhan untuk mengintegrasikan mekanisme pengukuran kemiripan teks ke dalam sistem ujian

secara menyeluruh. Berbeda dari penelitian terdahulu yang banyak membahas pengembangan atau perbandingan metode penilaian otomatis, penelitian ini berfokus pada implementasi mekanisme berbasis TF dan *Cosine Similarity* dalam aplikasi ujian *online* yang mencakup pengelolaan peserta didik, soal, ujian, pengerjaan ujian, koreksi otomatis, dan penyajian hasil. Pemilihan TF dan *Cosine Similarity* tidak dimaksudkan untuk menunjukkan keunggulan metode tersebut dibandingkan SBERT atau IndoBERT, melainkan sebagai pendekatan *lexical similarity* yang relatif sederhana dan ringan untuk diintegrasikan ke dalam aplikasi web. Kontribusi penelitian terletak pada aspek implementasi, yaitu mengintegrasikan pengolahan jawaban dan pengukuran kemiripan dengan keseluruhan proses ujian *online*, dengan menempatkan TF dan *Cosine Similarity* sebagai mekanisme pendukung otomatisasi koreksi.

Tujuan penelitian ini adalah mengimplementasikan TF dan *Cosine Similarity* pada aplikasi ujian *online* berbasis web untuk melakukan pengukuran kemiripan jawaban esai secara otomatis dan membantu penyajian hasil penilaian secara lebih terstruktur.

2. Metode Penelitian

Penelitian ini menggunakan metode pengembangan sistem dengan pendekatan *System Development Life Cycle* (SDLC) model Waterfall. Model ini digunakan karena tahapan pengembangan dilakukan secara berurutan, meliputi analisis kebutuhan, perancangan, implementasi, dan pengujian (Fitriani *et al.*, 2022; Saputri & Eriana, 2020). Objek penelitian berupa aplikasi ujian *online* esai koreksi otomatis berbasis web dengan data berupa kunci jawaban guru dan jawaban esai siswa. Pengumpulan data dilakukan melalui observasi, wawancara dengan guru, dan studi pustaka. Pendekatan pengembangan sistem secara bertahap juga banyak diterapkan pada pengembangan aplikasi berbasis web untuk memastikan kebutuhan pengguna diterjemahkan ke dalam fungsi sistem secara terstruktur (Al Hasri & Sudarmilah, 2021; Pane *et al.*, 2024). Model proses penelitian dapat dilihat pada Gambar 1.



Gambar 1. Model Proses Penelitian

2.1 Pengumpulan Data

Data dan informasi kebutuhan sistem diperoleh melalui observasi proses pembelajaran dan ujian, wawancara dengan guru terkait kebutuhan penilaian jawaban esai, dan studi pustaka tentang pengolahan teks, *Term Frequency*, *Cosine Similarity*, dan sistem ujian *online* (Arifuddin *et al.*, 2023). Data tersebut digunakan sebagai dasar dalam menentukan kebutuhan aplikasi dan mekanisme koreksi otomatis. Sumber dan teknik pengumpulan data disajikan pada Tabel 1.

Tabel 1. Sumber dan Teknik Pengumpulan Data

Sumber Data	Teknik	Data yang Diperoleh
Guru	Wawancara	Kebutuhan penilaian jawaban esai
Proses pembelajaran/ujian	Observasi	Alur pelaksanaan ujian dan koreksi
Literatur ilmiah	Studi pustaka	Teori TF, <i>Cosine Similarity</i> , sistem ujian <i>online</i>

Evaluasi metode dalam penelitian ini bersifat demonstratif karena data yang tersedia hanya mencakup satu pasangan kunci jawaban dan jawaban siswa untuk contoh perhitungan. Nilai *Cosine Similarity* yang diperoleh tidak digunakan sebagai ukuran akurasi keseluruhan sistem dan tidak digunakan untuk menyimpulkan kemampuan generalisasi metode. Pengujian *Black Box* dilakukan secara terpisah untuk mengevaluasi fungsi aplikasi. Perbandingan antara nilai sistem dan penilaian guru belum dapat dilakukan karena data penilaian guru dalam jumlah yang memadai tidak tersedia pada penelitian ini, sehingga validasi tersebut menjadi agenda penelitian lanjutan.

2.2 Pengembangan Sistem

Perancangan sistem mencakup rancangan proses, basis data menggunakan ERD, UML, dan antarmuka. Gambar 2 menampilkan *Use Case Diagram* yang menggambarkan hak akses setiap pengguna dalam sistem.



Gambar 2. Use Case Diagram

Admin dan siswa memiliki hak akses untuk *login* dan *logout* ke dalam sistem. Admin dapat mengelola data siswa, mata pelajaran, *quiz*, dan kelas, sedangkan siswa dapat melakukan ujian dan melihat hasil ujian. Tahap berikutnya adalah implementasi, yaitu membangun aplikasi berbasis web dan

mengintegrasikan metode koreksi otomatis, diikuti tahap pengujian untuk memastikan fungsi sistem berjalan sesuai kebutuhan.

2.3 Preprocessing dan Cosine Similarity

Jawaban siswa dan kunci jawaban diproses melalui tahap *case folding*, *tokenization*, *filtering*, dan *stemming* untuk menghasilkan teks yang lebih terstruktur sebelum dilakukan perhitungan kemiripan. Tahapan *preprocessing* tersebut relevan dengan penelitian penilaian jawaban otomatis yang menerapkan *casefolding*, *tokenizing*, dan *stopword removal* sebelum proses *similarity* (Pradani & Suadaa, 2023; Rangga Bakti *et al.*, 2025). Proses penilaian diawali dengan memasukkan kunci jawaban guru dan jawaban siswa, kemudian nilai kemiripan antara vektor kunci jawaban *A* dan jawaban siswa *B* dihitung menggunakan Persamaan (1).

$$\text{Cosine}(A, B) = \frac{A \cdot B}{\|A\| \|B\|} \quad (1)$$

Pendekatan TF dan *Cosine Similarity* telah digunakan dalam penilaian esai otomatis dan menunjukkan kemampuan untuk menghasilkan skor berdasarkan kemiripan jawaban (Pradani & Suadaa, 2023; Rangga Bakti *et al.*, 2025).

2.4 Implementasi dan Pengujian

Sistem dikembangkan sebagai aplikasi ujian *online* berbasis web yang melibatkan admin, guru, dan siswa. Fitur utama meliputi pengelolaan data siswa, mata pelajaran, kelas, soal, ujian, pengerjaan ujian, koreksi otomatis, dan hasil penilaian. Implementasi menggunakan Visual Studio Code, PHP, JavaScript, XAMPP, MySQL, dan Google Chrome sebagai lingkungan pengembangan dan pengujian. Pengujian sistem dilakukan menggunakan *Black Box Testing* untuk mengetahui kesesuaian fungsi aplikasi dengan kebutuhan yang telah ditentukan. Skenario pengujian meliputi *login*, *dashboard*, pengelolaan data siswa, ujian, dan hasil ujian (Zen *et al.*, 2024). Analisis dilakukan secara deskriptif dengan menghitung tingkat keberhasilan setiap skenario pengujian dan nilai kemiripan jawaban yang dihasilkan *Cosine Similarity*. Apabila tersedia skor penilaian guru, kesesuaian antara skor sistem dan skor manual dapat dianalisis menggunakan MAE, RMSE, atau korelasi untuk mengukur performa penilaian otomatis (Abidin *et al.*, 2025).

3. Hasil dan Pembahasan

3.1 Hasil

3.1.1 Hasil Preprocessing terhadap Kemiripan Jawaban

Hasil penelitian menunjukkan bahwa proses *preprocessing* berperan penting dalam membentuk *term* yang digunakan pada proses pengukuran kemiripan. *Stemming* menghasilkan bentuk dasar seperti *tuju*, sementara kata-kata yang sama pada kedua dokumen dipertahankan sebagai *term* untuk proses berikutnya. Contoh kunci jawaban dan jawaban siswa disajikan pada Tabel 2.

Tabel 2. Kunci Jawaban dan Jawaban Siswa

Kunci Jawaban	Jawaban Siswa
Wawancara adalah tanya jawab antara pewawancara dengan orang yang diwawancarai/narasumber dengan tujuan untuk memperoleh data untuk keperluan tertentu.	Wawancara ialah tanya jawab antara pewawancara dengan narasumber untuk memperoleh data tertentu.

Kedua teks diproses melalui *case folding*, *tokenization*, *filtering*, dan *stemming*. *Case folding* menyeragamkan karakter menjadi huruf kecil, *tokenization* memisahkan kalimat menjadi token, *filtering* menghilangkan kata yang tidak diperlukan, sedangkan *stemming* mengubah kata menjadi bentuk dasarnya. Tahapan

tersebut diterapkan pada kunci jawaban dan jawaban siswa sebelum proses pembobotan seperti yang disajikan pada Tabel 3.

Tabel 3. Hasil Preprocessing Kunci Jawaban dan Jawaban Siswa

Tahap	Kunci Jawaban	Jawaban Siswa
<i>Case Folding</i>	wawancara adalah tanya jawab antara pewawancara dengan orang yang diwawancarai narasumber dengan tujuan untuk memperoleh data untuk keperluan tertentu	wawancara ialah tanya jawab antara pewawancara dengan narasumber untuk memperoleh data tertentu
<i>Tokenization</i>	[wawancara, adalah, tanya, jawab, antara, pewawancara, dengan, orang, yang, diwawancarai, narasumber, dengan, tujuan, untuk, memperoleh, data, untuk, keperluan, tertentu]	[wawancara, ialah, tanya, jawab, antara, pewawancara, dengan, narasumber, untuk, memperoleh, data, tertentu]
<i>Filtering</i>	wawancara, pewawancara, orang, diwawancarai, narasumber, tujuan, memperoleh, data, keperluan	wawancara, pewawancara, narasumber, memperoleh, data
<i>Stemming</i>	wawancara, wawancara, orang, wawancara, narasumber, tuju, oleh, data, perlu	wawancara, wawancara, narasumber, oleh, data

Hasil *filtering* menyisakan kata-kata yang dianggap relevan, kemudian proses *stemming* mengubah kata berimbuhan menjadi bentuk dasarnya. Setelah *preprocessing*, dilakukan pembobotan berdasarkan *Term Frequency* (TF), yaitu jumlah kemunculan setiap *term* pada masing-masing dokumen. Semakin tinggi frekuensi kemunculan *term*, semakin besar kontribusinya terhadap proses pengukuran kemiripan.

3.1.2 Penerapan *Term Frequency* (TF)

Setelah dilakukan *preprocessing*, *term* direpresentasikan berdasarkan jumlah kemunculannya seperti yang disajikan pada Tabel 4.

Tabel 4. Hasil Representasi dan Kemiripan Jawaban

Term	Kunci Jawaban	Jawaban Siswa
Wawancara	3	2
Orang	1	0
Narasumber	1	1
Tuju	1	0
Oleh	1	1
Data	1	1
Perlu	1	0

Berdasarkan Persamaan (1), kunci jawaban direpresentasikan sebagai vektor *A* dan jawaban siswa direpresentasikan sebagai vektor *B*:

$$A = (3, 1, 1, 1, 1, 1, 1) \text{ dan } B = (2, 0, 1, 0, 1, 1, 0)$$

Perkalian titik kedua vektor:

$$A \cdot B = (3 \times 2) + (1 \times 0) + (1 \times 1) + (1 \times 0) + (1 \times 1) + (1 \times 1) + (1 \times 0) = 9$$

Panjang vektor kunci jawaban:

$$\| A \| = \sqrt{3^2 + 1^2 + 1^2 + 1^2 + 1^2 + 1^2 + 1^2} = \sqrt{15} = 3,873$$

Panjang vektor jawaban siswa:

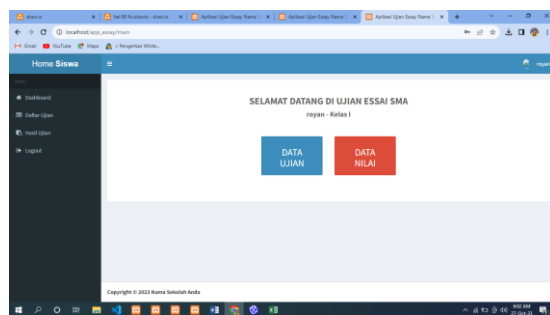
$$\| B \| = \sqrt{2^2 + 0^2 + 1^2 + 0^2 + 1^2 + 1^2 + 0^2} = \sqrt{7} = 2,646$$

$$\text{Cosine}(A, B) = \frac{9}{(3,873)(2,646)} = 0,878 = 87,8\%$$

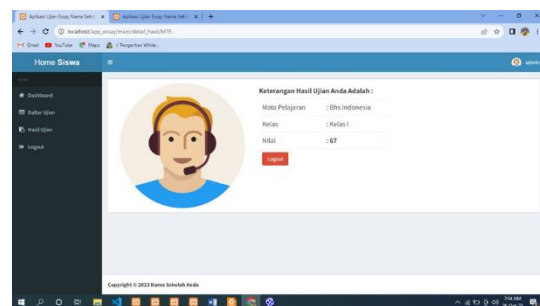
Hasil perhitungan menghasilkan nilai *Cosine Similarity* sebesar 0,878 atau 87,8% pada satu pasangan dokumen yang digunakan sebagai contoh pengujian. Nilai tersebut menunjukkan kemiripan tekstual dan tidak dapat diinterpretasikan sebagai akurasi sistem.

3.1.3 Implementasi Sistem

Penelitian ini menghasilkan aplikasi yang mengintegrasikan proses pencocokan jawaban ke dalam sistem ujian berbasis web. Sistem dirancang untuk mencocokkan jawaban siswa dengan kunci jawaban dan menghasilkan nilai secara otomatis.



Gambar 3. Halaman Dashboard



Gambar 4. Tampilan Hasil Penilaian Ujian Siswa

Halaman utama siswa menampilkan informasi ujian dan menyediakan akses ke menu Data Ujian serta Data Nilai. Melalui halaman ini, siswa dapat mengakses ujian yang tersedia dan melihat hasil penilaian setelah proses koreksi otomatis dilakukan oleh sistem. Tampilan dirancang sederhana agar siswa dapat mengakses fungsi utama sistem dengan mudah. Halaman hasil ujian menampilkan informasi mata pelajaran, kelas, dan nilai yang diperoleh siswa setelah proses koreksi otomatis (Gambar 4). Tampilan ini memberikan hasil penilaian secara langsung sehingga siswa dapat mengetahui hasil ujiannya tanpa menunggu proses koreksi manual oleh guru.

3.1.4 Hasil Pengujian Sistem

Pengujian dilakukan menggunakan metode *Black Box* untuk memastikan fungsi aplikasi berjalan sesuai dengan kebutuhan. Rencana pengujian mencakup fungsi *login*, *dashboard*, data siswa, ujian, hasil ujian, dan *logout*.

Tabel 6. Hasil Pengujian Fungsional

No.	Fitur yang Diuji	Hasil
1	<i>Login</i> Admin/Guru/Siswa	Berhasil
2	<i>Dashboard</i>	Berhasil
3	Data Siswa	Berhasil
4	Data Ujian	Berhasil

5	Hasil Ujian Siswa	Berhasil
6	Logout	Berhasil

Hasil pengujian menunjukkan bahwa fungsi utama aplikasi dapat dijalankan sesuai skenario pengujian berdasarkan masing-masing menu.

3.2 Pembahasan

Nilai *Cosine Similarity* sebesar 0,878 atau 87,8% menunjukkan bahwa pada pasangan dokumen contoh terdapat tingkat kemiripan tekstual yang relatif tinggi berdasarkan *term* yang digunakan. Nilai tersebut terutama dipengaruhi oleh *term* yang muncul pada kedua dokumen, yaitu *wawancara*, *narasumber*, *oleh*, dan *data*. *Term* wawancara memiliki kontribusi paling besar karena muncul tiga kali pada kunci jawaban dan dua kali pada jawaban siswa. Namun, nilai 87,8% tidak dapat disebut sebagai akurasi sistem karena nilai tersebut hanya menunjukkan kedekatan antara dua vektor berdasarkan representasi *term*. *Cosine Similarity* tidak secara langsung menentukan apakah isi jawaban benar secara substantif. Jawaban yang menggunakan banyak kata yang sama dapat memperoleh *similarity* tinggi meskipun argumentasinya belum tentu sama, sedangkan jawaban yang secara makna benar tetapi menggunakan sinonim atau struktur kalimat berbeda dapat memperoleh *similarity* lebih rendah. *Cosine Similarity* lebih tepat dipahami sebagai indikator kemiripan tekstual daripada ukuran langsung terhadap kebenaran substantif jawaban. Hal ini menjadi aspek penting dalam *Automated Essay Scoring* (AES). Dua jawaban dapat memiliki kata-kata yang sama tetapi belum tentu memiliki argumentasi atau konteks yang identik. Sebaliknya, dua jawaban dapat memiliki makna yang sama tetapi menggunakan kata yang berbeda. Temuan tersebut sejalan dengan perkembangan penelitian AES yang menunjukkan bahwa pendekatan berbasis *Cosine Similarity* dan representasi *term* memiliki keterbatasan dalam menangkap makna semantik.

Perbandingan penelitian ini dengan penelitian terdahulu disajikan pada Tabel 5.

Tabel 5. Perbandingan dengan Penelitian Terdahulu

Penelitian	Pendekatan	Evaluasi/Temuan	Posisi Penelitian
Lahitani (2022)	TF-IDF + <i>Cosine Similarity</i>	Penilaian esai multisoal	Menggunakan pembobotan <i>term</i> yang lebih kompleks
Chamidah et al. (2022)	SBERT + <i>Cosine Similarity</i>	MAE 0,26; Pearson 0,78	Lebih berorientasi pada kemiripan semantik
Kurniadi et al. (2023)	<i>Cosine Similarity</i>	<i>Similarity</i> rata-rata 85,4%; Kendall's W tinggi	Menunjukkan penggunaan <i>similarity</i> dan validasi terhadap instruktur
Pradani & Suadaa (2023)	TF-IDF + <i>Cosine</i> + IndoBERT	IndoBERT MAE 0,1285; RMSE 0,2001	Menggunakan representasi Transformer
Sitorus et al. (2026)	SBERT + Jaccard + Levenshtein	MAE 3,58; RMSE 4,48	Pendekatan <i>hybrid</i> semantik-leksikal
Penelitian ini	TF + <i>Cosine Similarity</i>	0,878 pada satu pasangan dokumen	Implementasi pendekatan leksikal pada aplikasi ujian web

Perbandingan tersebut menunjukkan bahwa penelitian ini berada pada pendekatan *lexical similarity*, yaitu mengukur kemiripan berdasarkan *term* yang terdapat pada dua dokumen. Pendekatan tersebut memiliki keunggulan dari sisi kesederhanaan komputasi dan kemudahan implementasi pada aplikasi web, tetapi memiliki keterbatasan ketika jawaban siswa menggunakan variasi bahasa yang berbeda dari kunci jawaban. Penelitian terdahulu yang menggunakan TF-IDF dan *Cosine Similarity* juga menunjukkan penggunaan pendekatan berbasis pembobotan *term* untuk AES, sementara penelitian yang menggunakan SBERT maupun IndoBERT menunjukkan bahwa representasi berbasis model bahasa dapat memberikan kemampuan yang lebih baik dalam menangkap kesamaan semantik.

4. Kesimpulan dan Rekomendasi

Penelitian ini menunjukkan bahwa penerapan *Term Frequency* (TF) dan *Cosine Similarity* dapat digunakan untuk mendukung proses koreksi otomatis jawaban esai berbasis web. Pendekatan tersebut mampu mengukur tingkat kemiripan antara jawaban siswa dan kunci jawaban berdasarkan kesamaan *term*, sehingga proses penilaian dapat dilakukan secara lebih terstruktur dan mengurangi ketergantungan pada koreksi manual. Pada contoh pengujian, diperoleh nilai kemiripan sebesar 87,8%, yang menunjukkan adanya tingkat kesamaan tekstual yang tinggi antara jawaban siswa dan kunci jawaban. Implementasi metode ke dalam aplikasi menghasilkan sistem yang dapat memproses jawaban dan menampilkan hasil penilaian kepada siswa. Pengujian fungsional juga menunjukkan bahwa fitur utama sistem dapat berjalan sesuai dengan skenario yang dirancang. Dengan demikian, tujuan penelitian untuk mengembangkan aplikasi yang membantu proses koreksi dan penyajian nilai esai secara otomatis telah tercapai.

Sebagai pengembangan selanjutnya, penelitian dapat dilakukan menggunakan jumlah data jawaban yang lebih besar dan membandingkan hasil penilaian sistem dengan penilaian guru menggunakan metrik evaluasi seperti MAE, RMSE, atau korelasi. Pengembangan metode berbasis kemiripan semantik juga dapat dipertimbangkan untuk mengatasi keterbatasan TF dan *Cosine Similarity* dalam mengenali jawaban yang memiliki makna sama tetapi menggunakan kata atau struktur kalimat yang berbeda.

5. Ucapan Terima Kasih

Terima kasih disampaikan kepada Universitas Pamulang atas dukungan dan fasilitas yang diberikan selama pelaksanaan penelitian. Terima kasih juga disampaikan kepada SMK Panti Karya 3 Gunung Sindur yang telah memberikan izin dalam proses pengumpulan dan analisis data sehingga penyusunan artikel ini dapat diselesaikan.

6. Daftar Pustaka

- Abidin, M. Z., Sudrajat, A. W., & Petrus, J. (2025). A systematic review of random forest and logistic regression algorithms for predicting student readiness in LSP-P1 competency certification at vocational high schools. *JuSiTik: Jurnal Sistem dan Teknologi Informasi Komunikasi*, 9(1), 76–89. <https://doi.org/10.32524/JUSITIK.V9I1.1735>
- Al Hasri, M. V., & Sudarmilah, E. (2021). Sistem informasi pelayanan administrasi kependudukan berbasis website Kelurahan Banaran. *MATRIK: Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, 20(2), 249–260. <https://doi.org/10.30812/MATRIK.V20I2.1056>
- Andhini, L., Maulana, R., & Pratama, N. R. (2025). Innovation in automated essay scoring for writing assessment: A case study at SMA Negeri 1 Genteng. *JINEA: Journal of Innovation in Education and Learning*, 1(3), 143–152. <https://doi.org/10.66031/JINEA.V1I3.35>
- Anggraeni, Y. S., Saputri, S. D., Azzahra, T., Verrel, A. G., & Tundjungsari, V. (2026). Implementasi content-based filtering menggunakan TF-IDF dan *Cosine Similarity* untuk rekomendasi buku akademik mahasiswa. *Jurnal Informatika: Jurnal Pengembangan IT*, 11(2), 213–225. <https://doi.org/10.30591/JPIT.V11I2.10055>

- Arifuddin, M. R., Rafiq, I. A., Mubarak, R., & Susilo, P. H. (2023). Sistem cerdas penilaian ujian essay menggunakan metode *Cosine Similarity*. *Generation Journal*, 7(1), 31–38. <https://doi.org/10.29407/GJ.V7I1.18318>
- Chamidah, N., Santoni, M. M., Irmanda, H. N., Astriratma, R., & Yulnelly, Y. (2022). Penilaian esai pendek otomatis berdasarkan similaritas semantik dengan SBERT. *Techno.Com*, 21(4), 732–743. <https://doi.org/10.33633/TC.V21I4.6758>
- Dillah, S., Syahyaningsih, L. T., Surianto, D. F., Budiarti, N. A. E., & Andayani, D. D. (2026). Integrating triplet loss with paraphrase-tuned SBERT for enhancing semantic meaning representation in Indonesian text. *International Journal of Fuzzy Logic and Intelligent Systems*, 26(1), 35–46. <https://doi.org/10.5391/IJFIS.2026.26.1.35>
- Fitriani, Y., Utami, S., & Junadi, B. (2022). Perancangan sistem informasi human capital management berbasis website. *Journal of Information System, Applied, Management, Accounting and Research*, 6(4), 792–803. <https://doi.org/10.52362/JISAMAR.V6I4.919>
- Gudya, S., Syahra, S., Sari, Y., & Suyanto, Y. (2022). The effect of text summarization in essay scoring (case study: Teach on e-learning). *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 16(1), 79–90. <https://doi.org/10.22146/ijccs.69906>
- Kurniadi, D., Gernowo, R., Surarso, B., Wibowo, A., & Warsito, B. (2023). Sistem penilaian jawaban singkat otomatis pada ujian online berbasis komputer menggunakan algoritma *Cosine Similarity*. *JEPIN*, 9(2), 316–322.
- Lahitani, A. R. (2022). Automated essay scoring menggunakan *Cosine Similarity* pada penilaian esai multi soal. *Jurnal Kajian Ilmiah*, 22(2), 107–118. <https://doi.org/10.31599/M8N1X493>
- 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), 477–485. <https://doi.org/10.31849/ZN.V6I2.20539>
- Permata, R. P., Suharto, R. N., Candra, L., & Julianty, A. (2025). Towards an automated essay evaluation system NLP based text embeddings and similarity metrics. *Digital Zone: Jurnal Teknologi Informasi dan Komunikasi*, 16(1), 37–46. <https://doi.org/10.31849/DIGITALZONE.V16I1.26541>
- Pradani, K. A., & Suadaa, L. H. (2023). Automated essay scoring menggunakan semantic textual similarity berbasis transformer untuk penilaian ujian esai. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 10(6), 1177–1184. <https://doi.org/10.25126/JTIK.2023107338>
- Rangga Bakti, I., Supriyanto, A., & Riki Mustafa, S. (2025). Sistem penilaian esai otomatis berbasis kecerdasan buatan menggunakan pendekatan text mining dan *Cosine Similarity*. *Riau Jurnal Teknik Informatika*, 4(3), 514–521. <https://doi.org/10.30606/RJTI.V4I3.4571>
- Rokhman, N., Maulan, P. A., & Wirahuda, N. A. (2025). Analisis penilaian esai secara otomatis menggunakan natural language processing (NLP) dan *Cosine Similarity*. *Go Infotech: Jurnal Ilmiah STMIK AUB*, 31(1), 41–52. <https://doi.org/10.36309/goi.v31i1.359>

- Saputra, R., Jayanta, & Pradana, M. G. (2024). Implementasi algoritma *Cosine Similarity* dan TF-IDF dalam menentukan rumpun jabatan. *Krea-TIF: Jurnal Teknik Informatika*, 12(1), 1–11. <https://doi.org/10.32832/kreatif.v12i1.15470>
- Saputri, G., & Eriana, E. S. (2020). Implementasi metode waterfall pada perancangan sistem informasi koperasi simpan pinjam berbasis web dan Android (studi kasus PT. PEB). *Jurnal Teknik Informatika*, 13(2), 133–146. <https://doi.org/10.15408/JTI.V13I2.17537>
- Sitorus, T. M., Ramadhan, E., & Kasyidi, F. (2026). Penilaian otomatis jawaban esai SMA menggunakan sentence-BERT dan hybrid Levenshtein-Jaccard dengan akurasi hybrid. *Jurnal Algoritma*, 23(1), 39–50. <https://doi.org/10.33364/ALGORITMA/V.23-1.3141>
- Suryani, L., & Edy, K. (2020). Pengembangan aplikasi "lost & found" berbasis Android dengan menggunakan metode term frequency-inverse document frequency (TF-IDF) dan *Cosine Similarity*. *Electro Luceat*, 6(2), 190–204. <https://doi.org/10.32531/JELEKN.V6I2.232>
- Zen, M., Irwan, I., Hafni, H., & Ananda, M. D. P. (2024). Implementasi dan pengujian menggunakan metode blackbox testing pada sistem informasi tracer study. *Bulletin of Computer Science Research*, 4(4), 327–340. <https://doi.org/10.47065/BULLETCR.V4I4.359>