

SKRIPSI
SISTEM PENILAIAN ESAI OTOMATIS MENGGUNAKAN *SEMANTIC*
TEXT SIMILARITY

Automated Essay Grading System Using Semantic Text Similarity



NURSYALINDA
D02 18 034

PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS SULAWESI BARAT
MAJENE
2025

LEMBAR PERSETUJUAN

SKRIPSI

**SISTEM PENILAIAN ESAI OTOMATIS MENGGUNAKAN *SEMANTIC
TEXT SIMILARITY***

Telah dipersiapkan dan disusun oleh :

NURSYALINDA

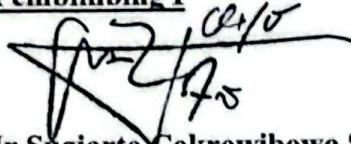
D0218034

Telah dipertahankan di depan Tim penguji

Pada tanggal 23 Juni 2025

Susunan Tim Penguji

Pembimbing I



Ir. Sugiarto Cokrowibowo S.Si, M.T

NIP. 199112272019031010

Pembimbing II



Nurhikma Arifin S.kom., M.T

NIP. 199304252022032011

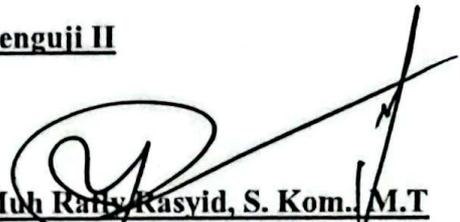
Penguji I



Arnita Irianti, S.Si., M.Si

NIP. 198708062018032001

Penguji II



Muh Rully Rasyid, S. Kom., M.T

NIP. 198808182022031006

Penguji III



Musawwir S.Pd., M.Pd

NIP. 19920909202406100

LEMBAR PENGESAHAN

SISTEM PENILAIAN ESAI OTOMATIS MENGGUNAKAN *SEMANTIC TEXT SIMILARITY*

SKRIPSI

Untuk memenuhi sebagian persyaratan
memperoleh gelar Sarjana Komputer

Disusun oleh:

NURSYALINDA
NIM. D0218034

Skripsi ini telah diuji dan dinyatakan lulus
pada Tanggal 23 Juni 2025
Telah diperiksa dan disetujui oleh:

Pembimbing I



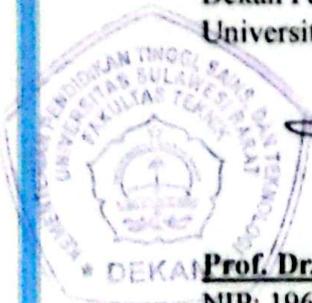
Ir. Sugiarto Cokrowibowo, S.Si., M.T
NIP.199112272019031010

Pembimbing II



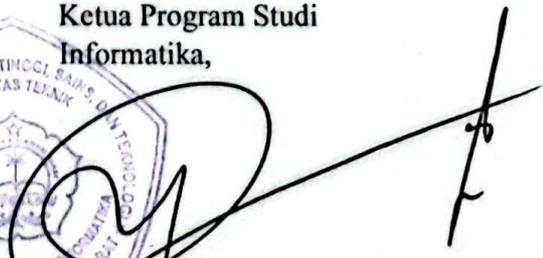
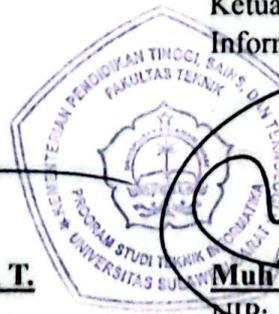
Nurhikma Arifin S.Kom., M.T
NIP.199304252022032011

Dekan Fakultas Teknik,
Universitas Sulawesi Barat



Prof. Dr. Ir. Hafsa Nirwana, M. T.
NIP: 196404051990032002

Ketua Program Studi
Informatika,



Muh Rafly Rasvid, S. Kom., M.T
NIP: 198808182022031006

PERNYATAAN ORISIONALITAS

Dengan ini saya menyatakan bahwa usulan skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjaan disuatu Perguruan Tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar Pustaka.

Apabila ternyata didalam naskah skripsi ini dapat dibuktikan terdapat unsur-unsur plagiasi, saya bersedia skripsi ini digugurkan dan gelar akademik yang telah saya peroleh (sarjana) dibatalkan, serta diproses sesuai dengan peraturan perundang-undangan yang berlaku (UU No.20 Tahun 2003,Pasal 25 Ayat 2 Dan Pasal 70)

Maiene. 23 Juni 2025



Nut syalinda
Nim:D0218034

ABSTRAK

Penilaian esai merupakan aspek krusial dalam pendidikan, namun seringkali menghadapi tantangan objektivitas dan efisiensi saat dilakukan secara manual. Penelitian ini mengusulkan pengembangan sistem penilaian esai otomatis menggunakan pendekatan Latent Semantic Analysis (LSA) yang terdiri dari Document Term Matrix (DTM) dan Singular Value Decomposition (SVD), serta metode Cosine Similarity. Tujuan utama penelitian ini adalah memvalidasi fungsionalitas sistem dalam memproses, menganalisis, dan memberikan penilaian *grade* otomatis. Data yang digunakan meliputi 10 jawaban esai responden dan satu set jawaban referensi. Proses pengujian dimulai dengan prapemrosesan data, dilanjutkan dengan pembentukan DTM untuk merepresentasikan esai dalam bentuk numerik. Selanjutnya, SVD diterapkan untuk mereduksi dimensi DTM, mengekstraksi konsep semantik laten, dan meningkatkan representasi dokumen dalam ruang topik. Tahap inti adalah perhitungan Cosine Similarity antara setiap jawaban responden dengan jawaban referensi, mengukur kemiripan semantik antar-dokumen. Berdasarkan nilai *cosine similarity* yang diperoleh, sistem secara fungsional mengonversi nilai tersebut menjadi *grade* A, B, C, D, atau E menggunakan ambang batas yang telah ditentukan. Hasil pengujian menunjukkan bahwa sistem berhasil melaksanakan setiap tahapan proses secara fungsional, dari prapemrosesan hingga penentuan *grade*. Misalnya, responden dengan *cosine similarity* tinggi (0.92915102 dan 0.99993911) secara konsisten mendapatkan *grade* tinggi, sementara responden dengan *cosine similarity* rendah (0.07950787) mendapatkan *grade* rendah. Ini membuktikan kapabilitas fungsional sistem dalam memberikan penilaian otomatis berdasarkan kedekatan semantik dengan jawaban referensi. Meskipun berfokus pada fungsionalitas, sistem ini menunjukkan potensi besar sebagai alat bantu penilaian yang objektif dan efisien.

Kata Kunci: Penilaian Esai Otomatis, *Latent Semantic Analysis* (LSA), *Document Term Matrix* (DTM), *Singular Value Decomposition* (SVD), *Cosine Similarity*, *Semantic Text Similarity*.

ABSTRACT

Essay grading is a crucial aspect of education, yet it often faces challenges regarding objectivity and efficiency when conducted manually. This research proposes the development of an automated essay grading system utilizing a Latent Semantic Analysis (LSA) approach, comprising Document Term Matrix (DTM) and Singular Value Decomposition (SVD), along with the Cosine Similarity method. The primary objective of this study is to validate the system's functionality in processing, analyzing, and assigning automated grades. The data used consists of 10 essay responses from various respondents and a set of reference answers. The testing process begins with data preprocessing, followed by the formation of a DTM to numerically represent the essays. Subsequently, SVD is applied to reduce the DTM's dimensionality, extract latent semantic concepts, and enhance document representation in a topic space. The core stage involves calculating the Cosine Similarity between each respondent's answer and the reference answer, measuring the semantic similarity between the documents. Based on the obtained *cosine similarity* values, the system functionally converts these values into grades A, B, C, D, or E using predefined thresholds. Test results indicate that the system successfully performs each process stage functionally, from preprocessing to grade determination. For instance, respondents with high *cosine similarity* (0.92915102 and 0.99993911) consistently received high grades, whereas respondents with low *cosine similarity* (0.07950787) received low grades. This demonstrates the system's functional capability in providing automated assessments based on semantic proximity to reference answers. Although focused on functionality, this system shows significant potential as an objective and efficient grading tool.

Keywords: *Automated Essay Grading, Latent Semantic Analysis (LSA), Document Term Matrix (DTM), Singular Value Decomposition (SVD), Cosine Similarity, Semantic Text Similarity.*

BAB I

PENDAHULUAN

1.1 Latar Belakang

Evaluasi dalam proses pembelajaran umumnya dilakukan melalui tiga bentuk utama ujian, yakni pilihan ganda, isian singkat, dan esai. Ujian esai yang menekankan pada kemampuan peserta didik dalam menyampaikan informasi secara verbal dan menguji kemampuan berpikir tingkat tinggi. Untuk menjawab tantangan penilaian esai secara manual yang membutuhkan waktu dan sumber daya besar, (Hamza, Sahriar, M. sarosa, 2013) mengembangkan sistem koreksi otomatis dengan algoritma *Rabin-Karp* dan *Hamming Distance*, yang efektif dalam mengenali kemiripan string antara jawaban siswa dan kunci jawaban.

Senada dengan itu, (Jakfar & Fergina, 2024) memanfaatkan metode *Cosine Similarity* dan teknik *stemming* untuk bahasa Indonesia guna menghitung kemiripan semantik jawaban esai, memungkinkan sistem penilaian otomatis yang efisien dan relevan dalam konteks pembelajaran daring. Kedua pendekatan ini menegaskan bahwa pemanfaatan teknologi dalam penilaian esai tidak hanya meningkatkan efisiensi, tetapi juga tetap mempertahankan kualitas evaluasi yang mendalam dan komprehensif.

Ujian esai lebih menuntut pemahaman yang lebih mendalam terhadap ilmu yang dipelajari karena formatnya memungkinkan pengungkapan ide secara komprehensif dan argumentatif, menjadikannya tolok ukur validitas pemahaman siswa dibandingkan format lain seperti isian atau pilihan ganda. (Jakfar & Fergina, 2024)

Menurut Kamus Besar Bahasa Indonesia, esai adalah “karangan prosa yang membahas suatu masalah secara sepintas lalu dari sudut pandang pribadi penulisnya”, yang berarti jenis evaluasi ini juga melatih siswa untuk menyampaikan informasi secara verbal dan reflektif, sesuai dengan karakteristik prosa esai.

Ujian esai memegang peranan penting dalam menilai pemahaman mahasiswa secara komprehensif karena menuntut kemampuan berpikir mendalam dan kemampuan mengorganisasi serta mengekspresikan gagasan

ilmiah—sebuah tolok ukur yang lebih representatif dibanding soal objektif biasa (Pradani & Suadaa, 2023). Esai menurut Kamus Besar Bahasa Indonesia (KBBI) adalah “*karangan prosa yang membahas suatu masalah secara sepintas lalu dari sudut pandang pribadi penulisnya*”.

Metode yang digunakan pada penelitian ini adalah *Latent Semantic Analysis* (LSA) untuk menganalisis struktur semantik suatu teks. LSA memanfaatkan model statistik matematis untuk mengkonversikan teks menjadi matriks-matriks yang diberi nilai pada masing-masing term untuk mencari kesamaan dengan term referensi. Dalam aplikasi ini, LSA digunakan untuk memberikan informasi tentang hubungan antar dokumen teks yang direpresentasikan menggunakan *term document matrix*. Aplikasi ini dapat memberikan nilai kemiripan pada setiap kata dalam dokumen. Meskipun LSA tidak memperhatikan tata letak kata pada dokumen, metode ini tetap efektif dalam mendeteksi kemiripan antar teks. (Karo Karo et al., 2020).

Dalam penelitian tentang evaluasi esai Bahasa Indonesia otomatis dengan analisis semantik laten penelitian ini mengimplementasikan sistem evaluasi otomatis esai dengan menggunakan Analisis Semantik Laten untuk mengukur tes esai. Metodologi penelitian ini meliputi pembersihan data, *case folding*, *tokenisasi*, *stop word*, konversi negasi, *stemming*, LSA, dan pengukuran kesamaan kosinus. Aplikasi ini ditujukan untuk esai dalam Bahasa Indonesia. (Amalia et al., 2019)

Penelitian ini bertujuan untuk mengevaluasi objektivitas penilaian esai otomatis. Serta mengukur sejauh mana kesamaan makna antara jawaban esai yang dihasilkan siswa dan jawaban ideal yang telah ditentukan oleh sistem. Untuk mencapai tujuan ini, akan dimanfaatkan berbagai model pendekatan semantik guna memvalidasi objektivitas nilai yang diberikan oleh sistem penilaian esai otomatis.

1.2 Rumusan Masalah

Adapun rumusan masalah pada penelitian ini adalah bagaimana implementasi *Latent Semantic Analysis* dalam penilaian esai otomatis menggunakan *semantic text similarity*.

1.3 Batasan Masalah

Agar penelitian ini menjadi terarah dan tidak menyimpang dari permasalahan yang ada penulis membatasi pokok permasalahan ini hanya pada :

1. Data yang digunakan pada penelitian ini hanya merujuk pada hasil jawaban responden.
2. Algoritma yang digunakan untuk menghitung data hasil ujian soal esai siswa menggunakan *Latent Semantic Analysis*.
3. Pengimplementasian diperuntukkan mengukur kesamaan makna antar jawaban siswa dan esai referensi (kunci jawaban).
4. Data yang akan dianalisis untuk pemberian skor *cosine similarity* dan *grade*.

1.4 Tujuan Penelitian

Untuk mengetahui bagaimana implementasi *Latent Semantic Analysis* dalam penilaian esai otomatis menggunakan semantik teks *similarity*.

1.5 Manfaat Penelitian

Manfaat penelitian yang diharapkan dari penelitian ini adalah:

1. Memberikan penilaian yang konsisten tanpa faktor subjektifitas penilaian manusia.
2. Menjadikan penelitian ini sebagai objek yang dapat berguna dimasa depan,serta mampu melahirkan ide-ide baru yang lebih mempermudah dalam pengambilan nilai secara tersistem dan objektif.

BAB II

TINJAUAN PUSTAKA

2.1 *Automated Essay Scoring*

Sebuah sistem penilaian jawaban uraian singkat (*essay*) secara otomatis. Dimana penilaian akan dilakukan secara komputerisasi dengan memperhitungkan nilai kedekatan teks antara jawaban peserta didik dengan kunci jawaban. Hasil dari sistem penilaian ini adalah berupa *grade* atau nilai kedekatan antara jawaban dengan kunci jawaban. Sebuah jawaban akan memiliki *grade* tinggi jika nilai kedekatan antara jawaban dengan kunci jawaban adalah tinggi, sebaliknya jika nilai kedekatan antara jawaban dengan kunci jawaban rendah maka *grade* yang dihasilkan juga rendah. (Hayatin & Jurusan, 2015)

2.2 *Semantic Text Similarity*

Semantic Text Similarity (STS) adalah tugas dalam pemrosesan bahasa alami yang melibatkan identifikasi keterkaitan semantik antara dua kalimat dalam bahasa yang sama atau berbeda. Mengidentifikasi kesamaan antara dua kalimat merupakan proses penting dalam banyak aplikasi, karena teks merupakan salah satu media komunikasi yang paling penting. STS memiliki banyak aplikasi di berbagai bidang seperti pencarian informasi, terjemahan mesin, deteksi plagiarisme, klasifikasi dokumen, pencarian semantik, dan sistem percakapan. (Al Sulaiman et al., 2022)

Dokumen teks merupakan rangkaian paragraf yang terdiri dari banyak kalimat, dimana pada kalimat itu sendiri merupakan rangkaian kata yang dapat mengungkapkan gagasan, pola pikir, serta perasaan. Kalimat merupakan satuan bahasa terkecil untuk mengungkap pikiran yang utuh, baik dalam lisan maupun tulisan. Pendekatan terhadap dokumen teks bahasa Indonesia dilakukan enkripsi untuk melihat struktur kalimatnya. (Detty Purnamasari, 2014).

2.3 *Text Mining*

Dalam artikel yang berjudul apa itu *text mining*? di jelaskan bahwa *text mining* merupakan sebuah informasi baru yang tidak diketahui sebelumnya oleh

komputer secara otomatis mengekstrak informasi dari sumber-sumber yang berbeda. Proses ini menggabungkan informasi yang telah berhasil diekstraksi dari berbagai sumber. Dengan tujuan untuk mengetahui secara lebih terkait informasi yang berguna dari beberapa dokumen dan mendukung proses *knowledge discovery* pada beberapa dokumen besar.

Berikut adalah beberapa algoritma yang umum menurut (Firdaus et al., 2021) digunakan dalam *text mining*:

1. Algoritma *text preprocessing*

Algoritma *text preprocessing* digunakan untuk memproses teks sebelum dilakukan analisis. Contoh algoritma *text preprocessing* yang digunakan dalam *text mining* adalah:

- a. *Tokenization*: algoritma ini menggunakan metode *tokenization* untuk membagi teks menjadi kata-kata yang lebih kecil.
 - b. *Stopword Removal*: algoritma ini menggunakan metode *stopwords* untuk menghapus kata-kata yang tidak relevan dalam teks.
2. Algoritma *Cosine Similarity*, untuk menentukan kesamaan antara teks esai dengan standar penilaian.
 3. Algoritma *Stemming*: mengoptimalkan proses penilaian esai dengan mengubah bentuk kata menjadi bentuk dasar.
 4. Algoritma *Text Summarization*

Algoritma *text summarization* digunakan untuk menggabungkan teks menjadi ringkasan yang lebih singkat. Contoh algoritma *text summarization* yang digunakan dalam *text mining* adalah:

- a. *Latent Semantic Analysis (LSA)*: algoritma ini menggunakan metode LSA untuk menggabungkan teks menjadi ringkasan yang lebih singkat.

2.4 Latent Semantic Analysis (LSA)

Latent Semantic Analysis (LSA) merupakan salah satu metode pengolahan bahasa alami yang banyak diterapkan dalam peringkasan teks berbahasa Indonesia dalam lima tahun terakhir. Riana & Nugraheni (2024) menggunakan kombinasi LSA dan TF-IDF untuk merangkum jurnal ilmiah, menunjukkan

bahwa kombinasi ini efektif menghasilkan ringkasan optimal pada kompresi 10–20 kalimat dengan nilai *f-measure* hingga 0,745. Sianturi, Ridho dan Santoso (2023) mengaplikasikan LSA pada artikel berita ekonomi, menggunakan variasi *compression rate* dan menemukan bahwa presisi terbaik mencapai 0,792 dengan akurasi hingga 0,902 pada kompresi 10%. Sementara itu, Nabilah Thahirah Kasim (2022) menerapkan LSA untuk peringkasan cerpen, memperoleh nilai *precision* sebesar 0,749 dan *f-measure* 0,714.

Ketiga studi tersebut menggunakan pendekatan yang serupa, yaitu *preprocessing (tokenizing, stopword removal, stemming)*, pembobotan TF-IDF, serta reduksi dimensi dengan SVD pada LSA. Hasil evaluasi dengan metrik seperti *precision, recall, f-measure, dan ROUGE* menunjukkan bahwa LSA mampu memberikan ringkasan yang relevan dan informatif, terutama ketika digunakan pada teks formal atau naratif dengan tingkat kompresi moderat (10–30%). Dengan demikian, LSA terbukti fleksibel dan efektif dalam mendukung pengembangan sistem peringkasan otomatis berbahasa Indonesia.

2.5 Cosine Similarity

Dalam berbagai studi pengukuran kemiripan teks berbahasa Indonesia, metode *cosine similarity* sering diterapkan untuk mendeteksi plagiarisme, kemiripan dokumen, dan sistem pencarian berbasis teks. Lumbansiantar et al., (2023) menunjukkan efektivitas metode ini dalam mendeteksi plagiarisme pada abstrak jurnal, dengan menghitung sudut vektor antar teks setelah tahap *preprocessing*. Hasil penelitian membuktikan bahwa *cosine similarity* mampu mengenali kesamaan struktur teks meskipun dikemas dalam kalimat berbeda, dengan nilai kemiripan yang diukur secara numerik, sehingga memudahkan analisis otomatis.

Pawestri & Suyanto, (2024) melakukan analisis komparatif menggunakan *cosine similarity, Jaccard, Euclidean*, dan Doc2Vec untuk teks dalam Bahasa Indonesia. Temuan mereka menyimpulkan bahwa *cosine similarity* memberikan performa terbaik dalam hal kecepatan dan akurasi (akurasi $\geq 98\%$, F1-score ≈ 0.89), terutama pada dataset berita *Google News* yang

terstruktur secara padat. Hasil ini mengonfirmasi bahwa representasi vektor sederhana dapat diandalkan pada pengolahan teks berbahasa alami dengan dimensi data sedang hingga tinggi.

Selain itu, implementasi *cosine similarity* juga diterapkan dalam domain skripsi, di mana kemiripan antar judul skripsi diukur antar 30 judul lama dan 5 judul baru. Sistem ini berhasil mendeteksi nilai kesamaan dengan rentang 0.1–0.49, yang menunjukkan metode ini dapat diadaptasi untuk kebutuhan sistem informasi akademik. Adaptasi dalam konteks pengolahan naskah skripsi ternyata bermanfaat dalam memitigasi plagiarisme potensi awal dengan mekanisme perbandingan otomatis berbasis teks. (Asriyani Arsad, et al., 2024).

2.6 Penelitian Terkait

Penelitian terkait bertujuan sebagai pendukung untuk melakukan penelitian yang digunakan sebagai referensi penulis. Adapun penelitian terkait dengan *Automates Essay Scoring System* menggunakan *Semantic text Similarity* diantaranya dapat dilihat pada tabel.

Tabel 2.1 Penelitian Terkait

No	Nama Peneliti	Judul Penelitian	Tujuan Penelitian	Hasil Penelitian & Kesimpulan	Persamaan	Perbedaan
1	A Amalia, D Gunawan, Y Fithri, I Aulia. (2019)	<i>Automated bahasa indonesia essay evaluation with latent semantic analysis</i>	Penelitian ini mengimplementasikan sebuah antarmuka <i>e-learning</i> untuk melakukan tes esai, menghasilkan sebuah sistem yang dapat menilai secara otomatis dengan menggunakan metode <i>Latent Semntic Analysis</i> (LSA).	Hasil dari penelitian ini akurasi sistem jika dibandingkan dengan penilaian manual oleh guru adalah 85% dengan harapan dapat membantu guru dalam melakukan penilaian yang lebih efisien.	Menggunakan metode LSA dengan memperhatikan tahapan <i>preprocessing</i> yang tepat untuk meningkatkan optimasi akurasi pada metode LSA.	Berfokus pada "pengembangan sistem penilaian esai otomatis" dan detail metode LSA yang digunakan.
2	Berlin Ong, Dali S. Naga, Viny Christanti M. (2020)	Perancangan Aplikasi Pendeteksi Kemiripan Teks Dengan	Memberikan informasi mengenai hubungan antar dokumen teks yang dipresentasikan	Menghasilkan dua buah modul utama yaitu modul <i>one to one</i> dan modul <i>one to money</i> .	Memberikan nilai pada teks dengan mengkonversikan teks menjadi matriks-	Mendeskripsikan fitur teknis spesifik (DTM untuk kemiripan kata, modul <i>one-to-</i>

		Menggunakan Metode Latent Semantic Analysis	menggunakan term dokumen matriks yang berguna untuk memberikan nilai kemiripan pada setiap kata dalam dokumen.	Yang berfungsi untuk membandingkan dua buah dokumen dalam bentuk teks atau lebih dari dua dokumen.	matriks yang diberi nilai pada masing-masing term untuk dicari kesamaan dengan <i>term</i> referensi. Dengan Tingkat akurasi yang didapatkan sebanyak 90% dengan menggunakan 10 data abstrak untuk percobaan pengujiannya.	<i>one/one-to-many</i>) dari sistem ini.
3	Meri Sri Wahyuni, Dedi Setiawan, Trinanda	Sistem Temu Kembali Informasi Dengan Latent Semantic Analisis Pada Kesamaan	Penelitian ini bertujuan untuk meminimalisir atau menghindari plagiat pada laporan tugas akhir mahasiswa dengan melakukan	Sistem pendeteksi plagiat yang diterapkan menggunakan metode <i>Latent Semantic Analysis (LSA)</i> dengan preprocessing dapat	Menggunakan perhitungan logaritma TF-IDF dan pembobotan <i>cosine similarity</i> dengan menganalisa setiap	Penelitian ini membahas terkait plagiasi menggunakan LSA untuk menemukan hubungan tersembunyi (<i>latent</i>)

	Syahputra (2021)	Tugas Akhir Mahasiswa	pendeteksian terhadap kesamaan dokumen menggunakan metode <i>Latent Semantic Analysis (LSA)</i> .	menghasilkan nilai <i>similarity</i> yang lebih akurat dan mendeteksi beberapa tipe plagiat dengan baik.	tahapan menggunakan metode <i>Latent Semantic Analysis</i> .	antara semua <i>term</i> (kata) yang memiliki kedekatan makna secara kontekstual.
4	Gita Indah Marthasari ¹ , Nur Hayatin, Maulidya Yuniarti. (2022)	<i>Content Classification Based-On Latent Semantic Analysis And Support Vector Machine (Lsa-Svm)</i>	Melakukan klasifikasi halaman <i>web</i> anak dengan menerapkan algoritma <i>Support Vector Machine</i> yang dikombinasikan dengan <i>Latent Semantic Analysis</i> .	Hasil pengujian klasifikasi halaman <i>web</i> anak menggunakan metode LSA-SVM didapatkan nilai akurasi sebesar 64% Dimana nilai ini lebih rendah dari klasifikasi menggunakan SVM.	Menggunakan metode LSA untuk membentuk sebuah matriks yang mepresentasikan hubungan antar <i>term</i> -dokumen yakni kata-kata dan dokumen-dokumen yang berhubungan dekat akan dihubungkan satu sama lain.	Mengklasifikasikan halaman web secara otomatis untuk membedakan web yang layak diakses oleh anak dan tidak dengan menggunakan algoritma SVM (<i>Support Vector Machine</i>)
5	Kamal Fauzan	Deteksi Plagiarisme Artikel Jurnal	Menganalisa kesamaan dokumen untuk deteksi	Dihasilkan rata-rata persentase akurasi	Alur penelitian ini menggunakan motode	diimplementasikan untuk mendeteksi

	Navaro, Septiyawan Rosetya Wardhana, Rinci Kembang Hapsari. (2023)	Menggunakan <i>Latent Semantic Analysis</i> (Lsa)	plagiarism jurnal menggunakan LSA berbasis web serta mengukur kinerja dari metode <i>Latent Semantic Analysis</i> (LSA).	sebesar 87.88% yang dinilai cukup besar sehingga sistem yang dibuat dinilai cukup bagus.	LSA untuk memberikan nilai dalam teks dengan cara mengubah teks menjadi sebuah matriks yang memberikan nilai pada setiap istilah guna mencari persamaan denga istilah acuannya. LSA menerapkan teknik matematika aljabar linier, khususnya SVD (<i>Singular Value Decomposition</i>).	plagiarisme artikel jurnal.
6	Manat Hendry Fernando	Peringkasan Teks Otomatis Menggunakan	Untuk membantu mendapatkan intisari dari dokumen. Yang	menghasilkan <i>precison</i> dan akurasi terbaik pada saat <i>compression rate</i>	menggunakan metode <i>Latent Semantic Analysis</i> yang	Melakukan peringkasan teks pada sebuah artikel berita

	Sianturi, Achmad Ridok, Edy Santoso. (2023)	Metode <i>Latent Semantic Analysis</i> Pada Artikel Berita Ekonomi Berbahasa Indonesia	dipenelitian ini memaparkan peringkasan artikel berita ekonomi berbahasa Indonesia.	10% dengan nilai 0.7416 dan akurasi 0.909853. Sedangkan untuk <i>recall</i> , dan <i>f-measure</i> terbaik dihasilkan pada saat <i>compression rate</i> 50% dengan nilai 0.25201 dan 0.284788.	menggunakan pendekatan aljabar <i>linear singular value decomposition</i> (SVD) dengan membentuk matriks representasi dari asosiasi <i>term</i> yang merupakan kata-kata pada dokumen yang berhubungan erat dari proses perhitungan TF-IDF	ekonomi menggunakan <i>Latent Semantic Analysis</i> .
--	---	---	---	---	--	--

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil pengujian dan pembahasan yang telah dilakukan dalam penelitian ini, dapat ditarik beberapa kesimpulan terkait pengembangan sistem penilaian esai otomatis menggunakan algoritma Latent Semantic Analysis (LSA):

1. Sistem penilaian esai otomatis yang dikembangkan telah berhasil menunjukkan kapabilitas fungsionalnya dalam mengolah jawaban esai responden. Setiap tahapan proses, mulai dari prapemrosesan teks, pembentukan *Document Term Matrix* (DTM), reduksi dimensi menggunakan *Singular Value Decomposition* (SVD), hingga perhitungan *Cosine Similarity*, telah berjalan dengan baik sesuai rancangan. Sistem mampu mengubah esai mentah menjadi representasi numerik yang dapat dianalisis secara komputasi.
2. Implementasi LSA terbukti efektif dalam menangkap kemiripan semantik antara jawaban responden dan jawaban referensi. Nilai *Cosine Similarity* yang dihasilkan secara intuitif mencerminkan kedekatan atau perbedaan konten antara esai yang dinilai dengan standar.
3. Sistem dapat diimplementasikan menggunakan *google form* secara otomatis dengan menghasilkan file *xlsx*.
4. Sistem mampu mengonversi nilai *Cosine Similarity* menjadi *grade* kategorikal (A, B, C, D, E) berdasarkan ambang batas yang telah ditetapkan. Fungsionalitas ini memudahkan interpretasi hasil penilaian oleh pengguna.

5.2 Saran

Berdasarkan hasil penelitian dan keterbatasan yang ada, beberapa saran diajukan untuk pengembangan lebih lanjut:

1. Meskipun fokus penelitian ini adalah fungsionalitas, pada pengembangan selanjutnya disarankan untuk melakukan pengujian akurasi yang lebih mendalam. Hal ini dapat dicapai dengan membandingkan *grade* otomatis

yang dihasilkan sistem dengan penilaian manual oleh penilai manusia yang kompeten, menggunakan metrik seperti *Quadratic Weighted Kappa* (QWK) atau korelasi Pearson.

2. Eksplorasi teknik prapemrosesan yang lebih canggih (misalnya, *lemmatization* yang lebih presisi, penanganan sinonim, atau *part-of-speech tagging*) dapat meningkatkan kualitas representasi DTM dan SVD, berpotensi memperbaiki akurasi semantik.
3. Penelitian ini menggunakan 2 dimensi SVD. Eksperimen dengan jumlah dimensi laten yang berbeda (misalnya, 5, 10, atau lebih) dapat dilakukan untuk mencari konfigurasi optimal yang paling efektif dalam menangkap konsep semantik dan meminimalkan *noise*, yang mungkin berdampak pada akurasi penilaian.
4. Selain kemiripan semantik, sistem dapat diperkaya dengan mempertimbangkan fitur-fitur lain yang memengaruhi kualitas esai, seperti tata bahasa, ejaan, struktur kalimat, atau koherensi argumen. Hal ini dapat diintegrasikan melalui pendekatan *multi-feature* untuk penilaian yang lebih komprehensif.

DAFTAR PUSTAKA

- Al Sulaiman, M., Moussa, A. M., Abdou, S., Elgibreen, H., Faisal, M., & Rashwan, M. (2022). Semantic textual similarity for modern standard and dialectal Arabic using transfer learning. *PLoS ONE*, *17*(8 August), 1–14. <https://doi.org/10.1371/journal.pone.0272991>
- Amalia, A., Gunawan, D., Fithri, Y., & Aulia, I. (2019). Automated Bahasa Indonesia essay evaluation with latent semantic analysis. *Journal of Physics: Conference Series*, *1235*(1). <https://doi.org/10.1088/1742-6596/1235/1/012100>
- Ariantini, D. A. R., Lumenta, A. S. M., & Jacobus, A. (2016). Pengukuran Kemiripan Dokumen Teks Bahasa Indonesia Menggunakan Metode Cosine Similarity. *Jurnal Teknik Informatika*, *9*(1), 1–8. <https://doi.org/10.35793/jti.9.1.2016.13752>
- Firdaus, A., Firdaus, W. I., Studi, P., Informatika, T., Digital, M., & Sriwijaya, P. N. (2021). *Text Mining*. *13*(1), 66–78.
- Hamza, Sahriar, M. sarosa, purnomo budi santoso. (2013). 61255-ID-sistem-koreksi-soal-essay-otomatis-dengan metode RABIN KARB. *Eeccis*, *7*, 153–158.
- Hayatin, N., & Jurusan, K. (2015). Sistem Penilaian Jawaban Essay Otomatis Berdasarkan Nilai Kedekatan Kalimat. *Seminar Teknologi Dan Rekayasa (SENTRA)*, 978–979.
- Jakfar, A. E., & Fergina, A. (2024). *ANALISIS UJIAN ESSAY MENGGUNAKAN METODE*. *6*(2), 349–357.
- Karo Karo, B. O., Naga, D. S., & Mawardi, V. C. (2020). Perancangan Aplikasi Pendeteksi Kemiripan Teks Dengan Menggunakan Metode Latent Semantic Analysis. *Computatio: Journal of Computer Science and Information Systems*, *4*(1), 1. <https://doi.org/10.24912/computatio.v4i1.7191>
- Lumbansiantar, S., Dwiasnati, S., & Fatonah, N. S. (2023). Penerapan Metode Cosine Similarity Dalam Mendeteksi Plagiarisme Pada Jurnal. *Format: Jurnal Ilmiah Teknik Informatika*, *12*(2), 142. <https://doi.org/10.22441/format.2023.v12.i2.007>
- Nabilah Thahirah Kasim. (2022). *Peringkasan Teks Bahasa Indonesia Pada Cerpen Menggunakan Metode Latent Semantic Analysis (Lsa)*.
- Pawestri, S., & Suyanto, Y. (2024). Analisis Perbandingan Metode Similarity untuk Kemiripan Dokumen Bahasa Indonesia pada Deteksi Kemiripan Teks Bahasa Indonesia. *Jurnal Media Informatika Budidarma*, *8*(3), 1440. <https://doi.org/10.30865/mib.v8i3.7648>
- 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/jtiik.1067338>

- Riana, D., & Nugraheni, M. (2024). *PERINGKASAN TEKS BERBAHASA INDONESIA MENGGUNAKAN TEKNIK EKSTRAKSI DENGAN ALGORITMA LATENT SEMANTIC ANALYSIS (LSA) DENGAN VARIASI TF-IDF UNTUK PERINGKASAN*. 8(1), 94–101.
- Sianturi, M. H. F., Ridok, A., & Santoso, E. (2023). Peringkasan Teks Otomatis menggunakan Metode Latent Semantic Analysis pada Artikel Berita Ekonomi berbahasa Indonesia. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 7(5), 2589–2596. <http://j-ptiik.ub.ac.id>
- Teknik, F. (2024). Penerapan Teks Mining Dan Cosine Similarity Untuk Menentukan Kesamaan Dokumen Skripsi Application of Text Mining and Cosine Similarity To Determine the Similarity of Thesis Documents. *IJIS Indonesian Journal on Information System*, 9(April 2024), 99–109.