

Analisis Penilaian Esai Secara Otomatis Menggunakan *Natural Language Processing (NLP)* dan *Cosine Similarity*

Nur Rokhman^{*1}, Puteri Anindya Maulan², Novriansyah Alief Wirahuda³

^{*1,2} D4 Akuntansi Perpajakan Universitas Sain dan Teknologi Komputer Semarang, Indonesia

³ S1 Sistem Informasi Universitas Sain dan Teknologi Komputer Semarang, Indonesia

e-mail: ^{*1} nurrokhman087@gmail.com, ² putri@stekom.ac.id, ³ blenyixgaptech@gmail.com

Abstrak

Penilaian esai merupakan salah satu bentuk evaluasi yang efektif dalam mengukur pemahaman konseptual dan kemampuan berpikir kritis mahasiswa. Namun, proses penilaian manual terhadap jawaban esai seringkali menghadapi kendala berupa waktu koreksi yang lama, subjektivitas penilai, serta inkonsistensi antar dosen. Penelitian ini bertujuan untuk mengembangkan dan menguji efektivitas sistem penilaian otomatis berbasis *Natural Language Processing (NLP)* dan algoritma *Cosine Similarity* terhadap jawaban esai mahasiswa. Metode penelitian menggunakan pendekatan eksperimen kuantitatif dengan data uji berasal dari mata kuliah *Struktur Data*. Proses meliputi pra-pemrosesan teks (*tokenisasi*, *stopword removal*, *stemming*), representasi dokumen menggunakan *TF-IDF*, serta pengukuran kemiripan semantik antara jawaban mahasiswa dan kunci jawaban. Hasil pengujian menunjukkan bahwa sistem memiliki rata-rata persentase kemiripan sebesar 88,2% dengan nilai *Mean Absolute Error (MAE)* sebesar 2, yang mencerminkan performa yang baik dan stabil. Selain itu, waktu penilaian dapat dipersingkat hingga lebih dari 60% dibandingkan metode manual. Temuan ini menunjukkan bahwa pendekatan *NLP* dan *Cosine Similarity* layak diterapkan sebagai sistem pendukung dalam evaluasi akademik berbasis teks di lingkungan pendidikan tinggi.

Kata kunci— penilaian esai, *natural language processing*, *cosine similarity*, *TF-IDF*, evaluasi otomatis, sistem informasi akademik

Abstract

Essay assessment is a form of evaluation that is effective in measuring students' conceptual understanding and critical thinking skills. However, the manual assessment process for essay answers often faces obstacles in the form of long correction times, assessor subjectivity, and inconsistencies between lecturers. This study aims to develop and test the effectiveness of an automated assessment system based on *Natural Language Processing (NLP)* and the *Cosine Similarity* algorithm on students' essay answers. The research method uses a quantitative experimental approach with test data from the *Data Structure* course. The process includes pre-processing of the text (*tokenization*, *stopword removal*, *stemming*), representation of documents using *TF-IDF*, and measurement of semantic similarity between students' answers and answer keys. The test results showed that the system had an average percentage of similarity of 88.2% with a *Mean Absolute Error (MAE)* value of 2, which reflected good and stable performance. In addition, the assessment time can be shortened by more than 60% compared to the manual method. These findings suggest that *NLP* and *Cosine Similarity* approaches are feasible to be applied as a support system in text-based academic evaluation in higher education settings

Keywords— essay assessment, *natural language processing*, *cosine similarity*, *TF-IDF*, automated evaluation, academic information systems

1. PENDAHULUAN

1.1 Latar Belakang

Evaluasi hasil belajar merupakan elemen krusial dalam sistem pendidikan tinggi yang berperan dalam menilai sejauh mana mahasiswa memahami materi perkuliahan yang telah disampaikan [9]. Umumnya, proses evaluasi akademik dilakukan melalui Ujian Tengah Semester (UTS) dan Ujian Akhir Semester (UAS), dengan salah satu metode yang banyak digunakan adalah ujian berbentuk esai. Ujian esai memiliki beberapa keunggulan dibandingkan jenis evaluasi lainnya. Pertama, bentuk soal esai memungkinkan pengujian pemahaman konseptual secara mendalam, karena mahasiswa diharuskan menjelaskan pemikiran mereka dengan bahasa sendiri, bukan sekadar memilih jawaban [11]. Kedua, ujian esai merangsang kemampuan berpikir tingkat tinggi seperti analisis, sintesis, dan evaluasi. Ketiga, fleksibilitas dalam menjawab memungkinkan mahasiswa mengekspresikan pemahamannya melalui berbagai pendekatan yang sah [7]. Meskipun demikian, metode esai juga menghadirkan tantangan signifikan, terutama dalam proses penilaiannya. Penilaian secara manual rentan terhadap subjektivitas karena melibatkan interpretasi dosen terhadap jawaban mahasiswa, yang bisa berbeda-beda [4]. Selain itu, penilaian manual membutuhkan waktu yang tidak sedikit, terutama pada mata kuliah dengan jumlah peserta yang besar. Faktor lain yang menjadi perhatian adalah ketidakkonsistenan skor, terutama bila penilaian dilakukan oleh lebih dari satu pengujii [8].

Seiring dengan perkembangan teknologi informasi dan kecerdasan buatan (Artificial Intelligence/AI), pendekatan baru untuk mengotomatisasi penilaian esai mulai dikembangkan. Salah satu solusi yang potensial adalah penerapan Natural Language Processing (NLP) dan algoritma Cosine Similarity dalam sistem evaluasi akademik. Pendekatan ini diharapkan mampu memberikan alternatif yang lebih efisien, objektif, dan konsisten dalam proses penilaian.

1.2 Urgensi Penelitian

Kebutuhan akan sistem evaluasi yang cepat dan adil mendorong munculnya inovasi berbasis AI, khususnya dalam konteks penilaian otomatis. Teknologi NLP telah banyak digunakan dalam analisis bahasa alami, termasuk dalam konteks pendidikan untuk menilai esai dan tanggapan terbuka [12]. Penerapan NLP dalam pendidikan dapat mengurangi beban kerja dosen serta meningkatkan akurasi dan konsistensi penilaian. Salah satu teknik NLP yang umum digunakan dalam mengukur kesamaan teks adalah algoritma Cosine Similarity. Algoritma ini bekerja dengan membandingkan representasi vektor dari dua teks untuk menghitung tingkat kemiripan antar dokumen [5]. Dalam konteks penilaian esai, metode ini berfungsi membandingkan jawaban mahasiswa dengan kunci jawaban yang telah disiapkan, sehingga sistem dapat memberikan skor secara otomatis berdasarkan derajat kemiripan yang dihitung secara matematis.

Dengan demikian, integrasi NLP dan Cosine Similarity dalam sistem penilaian esai tidak hanya menjadi solusi atas tantangan yang ada, tetapi juga menjadi langkah strategis menuju transformasi digital dalam sistem informasi akademik. Perbandingan antara metode penilaian manual dan penilaian otomatis berbasis NLP ditampilkan pada Tabel 1 berikut:

Tabel 1. Perbandingan Penilaian Manual dan Otomatis Berbasis NLP

Aspek	Penilaian Manual	Penilaian Otomatis Berbasis NLP
Kecepatan	Lambat (butuh waktu berhari-hari untuk menilai banyak jawaban)	Cepat (penilaian bisa dilakukan dalam hitungan detik)
Objektivitas	Cenderung subjektif (dipengaruhi oleh interpretasi dosen)	Objektif (berdasarkan perhitungan algoritma)
Konsistensi Penilaian	Rentan terhadap perbedaan antar dosen	Konsisten karena berdasarkan analisis teks yang terukur
Kemampuan Menilai Argumentasi	Baik, karena manusia bisa memahami konteks	Terbatas, karena hanya mengukur kesamaan kata/frase
Skalabilitas	Sulit diterapkan untuk jumlah mahasiswa yang banyak	Mudah diterapkan untuk ribuan jawaban esai

Berdasarkan Tabel 1, metode penilaian otomatis berbasis NLP memiliki keunggulan dalam kecepatan, objektivitas, dan konsistensi, meskipun masih memiliki keterbatasan dalam menilai kualitas argumentasi mahasiswa.

1.3 Tujuan Penelitian

Penelitian ini disusun untuk menjawab permasalahan yang muncul dalam proses penilaian ujian esai secara konvensional yang selama ini masih bersifat manual, memakan waktu, dan rentan terhadap subjektivitas. Secara khusus, penelitian ini bertujuan untuk:

1. Merancang, mengembangkan Sistem Penilaian Esai Otomatis Berbasis NLP dan Cosine Similarity

Tujuan utama penelitian ini adalah membangun sistem informasi yang mampu mengevaluasi jawaban esai mahasiswa secara otomatis. Sistem ini dirancang dengan pendekatan Natural Language Processing (NLP) dan algoritma Cosine Similarity untuk mengukur tingkat kesamaan semantik antara teks jawaban mahasiswa dan kunci jawaban. Dengan pendekatan ini, sistem dapat memberi penilaian yang objektif, cepat, dan konsisten.

2. Mengevaluasi Efektivitas Teknik Natural Language Processing dalam Penilaian Esai

Penelitian ini bertujuan untuk menganalisis kontribusi berbagai teknik NLP seperti tokenisasi, stopword removal, stemming, dan representasi Term Frequency-Inverse Document Frequency (TF-IDF) dalam meningkatkan akurasi sistem penilaian. Evaluasi ini dilakukan untuk mengetahui sejauh mana NLP dapat memahami konteks dan struktur kalimat dalam jawaban esai secara komputasional.

3. Mengukur Efisiensi Waktu dan Akurasi Hasil Penilaian Otomatis

Efisiensi merupakan salah satu faktor penting dalam implementasi sistem otomatis. Oleh karena itu, penelitian ini ingin mengetahui seberapa besar penghematan waktu yang dicapai oleh sistem dalam proses evaluasi dibandingkan dengan metode manual, serta tingkat akurasi hasil yang diperoleh.

4. Membandingkan Hasil Penilaian Otomatis dengan Penilaian Manual oleh Pengajar

Validasi sistem dilakukan dengan membandingkan hasil skor dari sistem otomatis dengan skor yang diberikan oleh dosen pengampu secara manual. Tujuan dari perbandingan ini adalah untuk mengetahui apakah sistem dapat berfungsi sebagai alat bantu atau bahkan sebagai alternatif dalam proses evaluasi pembelajaran.

5. Menyediakan Solusi Teknologi terhadap Permasalahan Penilaian Ujian Esai Skala Besar

Penelitian ini diharapkan dapat menawarkan solusi yang praktis dan terintegrasi dalam sistem informasi akademik untuk menjawab tantangan dalam penilaian esai berskala besar, seperti keterbatasan waktu, beban kerja dosen, dan ketidakkonsistenan dalam pemberian nilai.

1.4 Konteks Permasalahan

Penilaian merupakan aspek penting dalam proses pembelajaran di perguruan tinggi, terutama sebagai indikator pencapaian kompetensi mahasiswa. Salah satu bentuk evaluasi yang dianggap mampu mengukur kemampuan berpikir kritis, kedalaman pemahaman, serta keterampilan analitis mahasiswa adalah ujian esai. Namun demikian, karakteristik jawaban esai yang bersifat terbuka menghadirkan tantangan signifikan dalam hal objektivitas, efisiensi waktu, dan konsistensi nilai antar penilai.

Kemajuan dalam bidang kecerdasan buatan, khususnya Natural Language Processing (NLP), membuka peluang untuk mengatasi kendala tersebut. NLP memungkinkan komputer memahami struktur dan makna bahasa alami, sehingga cocok digunakan dalam proses analisis teks. Dengan memanfaatkan algoritma Cosine Similarity, sistem dapat menghitung tingkat kemiripan antara teks jawaban mahasiswa dan kunci jawaban berbasis representasi vektor. Hal ini memungkinkan proses penilaian dilakukan secara otomatis dengan mempertimbangkan aspek semantik teks.

Melalui pendekatan ini, penelitian diarahkan pada pengembangan sistem penilaian esai otomatis berbasis NLP dan Cosine Similarity yang mampu meningkatkan efisiensi, objektivitas, serta akurasi penilaian dalam skala besar. Sistem ini diharapkan dapat menjadi bagian dari inovasi teknologi dalam sistem informasi akademik yang mendukung transformasi digital dalam dunia pendidikan tinggi.

1.5 Ruang Lingkup Penelitian

Agar fokus dan arah penelitian ini tetap berada dalam koridor kajian yang relevan, ruang lingkup ditetapkan secara terperinci. Pembatasan ini bertujuan untuk memastikan bahwa seluruh proses penelitian selaras dengan tujuan yang hendak dicapai serta dapat menghasilkan kontribusi ilmiah yang terukur.

1. Jenis dan Sumber Data

Data yang digunakan dalam penelitian ini berasal dari kumpulan jawaban esai mahasiswa pada mata kuliah Struktur Data dalam program studi Sistem Informasi di salah satu perguruan tinggi. Jawaban diperoleh dari hasil Ujian Akhir Semester dan telah tersedia dalam format digital (teks). Kunci jawaban disusun oleh dosen pengampu dan dijadikan acuan utama dalam proses evaluasi otomatis. Seluruh data bersifat anonim untuk menjaga kerahasiaan identitas responden.

2. Metode dan Teknik Pengolahan Teks

Penelitian ini menerapkan pendekatan Natural Language Processing (NLP) untuk mengolah data teks secara sistematis. Proses awal meliputi tahapan tokenisasi, konversi ke huruf kecil (lowercasing), penghapusan kata-kata umum yang tidak memiliki makna khusus (stopword removal), serta stemming dan lemmatization untuk mengembalikan kata ke bentuk dasarnya. Dokumen hasil pra-pemrosesan selanjutnya dikonversi ke dalam representasi numerik menggunakan metode Term Frequency–Inverse Document Frequency (TF-IDF). Pengukuran tingkat kemiripan antara jawaban mahasiswa dan kunci jawaban dilakukan menggunakan algoritma Cosine Similarity, yang bekerja berdasarkan sudut antar vektor dalam ruang multidimensi representasi dokumen.

3. Parameter yang Diukur

Fokus utama sistem yang dikembangkan adalah mengukur kesamaan semantik antara jawaban esai mahasiswa dengan kunci jawaban. Hasil pengukuran ditampilkan dalam bentuk nilai persentase kemiripan, mulai dari 0% (tidak mirip sama sekali) hingga 100% (sangat mirip). Persentase ini digunakan sebagai dasar perhitungan skor akhir, dengan asumsi bahwa semakin tinggi kemiripan, semakin sesuai jawaban mahasiswa dengan ekspektasi pengajar.

4. Batasan Penelitian

Penelitian ini memiliki sejumlah keterbatasan yang perlu dicatat. Pertama, sistem hanya melakukan penilaian berdasarkan kemiripan statistik antar teks, tanpa mempertimbangkan aspek kualitas isi seperti logika argumentasi, struktur kalimat, keaslian ide, atau gaya

penulisan. Kedua, pendekatan yang digunakan masih berbasis metode NLP konvensional dan belum mengintegrasikan model pembelajaran mendalam (deep learning) seperti BERT, RoBERTa, atau Word2Vec yang mampu menangkap makna kontekstual dan relasi sinonim secara lebih kompleks.

Pengembangan sistem lanjutan diperlukan untuk mengadopsi teknologi NLP yang lebih canggih agar hasil penilaian lebih menyeluruh dan mendekati pemahaman manusia terhadap teks.

1.6 Kontribusi Penelitian

Penelitian ini diharapkan mampu memberikan kontribusi yang bermakna, baik dari sisi teoretis maupun praktis, khususnya dalam pengembangan sistem informasi untuk pendidikan dan teknologi evaluasi akademik. Manfaat yang dapat diidentifikasi berdasarkan lingkup penggunaannya dijelaskan sebagai berikut:

1. Manfaat bagi Dosen dan Tenaga Pengajar

Implementasi sistem penilaian otomatis berbasis Natural Language Processing (NLP) dan algoritma Cosine Similarity memungkinkan pengajar menghindari proses koreksi manual terhadap ratusan lembar jawaban esai. Otomatisasi ini dapat menghemat waktu dan tenaga sekaligus meminimalkan kelelahan serta ketidakkonsistenan dalam pemberian skor. Peningkatan Fokus terhadap Aktivitas Pembelajaran, dengan berkurangnya beban administratif, dosen memiliki lebih banyak waktu untuk dialokasikan pada pengembangan materi ajar, kegiatan bimbingan akademik, serta penelitian ilmiah. Hal ini mendukung peran dosen dalam menciptakan proses pembelajaran yang lebih produktif dan bermakna.

2. Manfaat bagi Institusi Pendidikan

Sistem yang dikembangkan dapat diintegrasikan ke dalam sistem informasi akademik atau Learning Management System (LMS) institusi untuk mempercepat proses evaluasi dalam skala besar tanpa mengorbankan akurasi maupun objektivitas hasil penilaian. Penelitian ini turut mendukung visi digitalisasi pendidikan tinggi, terutama dalam konteks kebijakan Kampus Merdeka, dengan memperkenalkan pendekatan teknologi yang inovatif untuk mendukung proses evaluasi berbasis teks terbuka secara sistematis dan efisien.

3. Manfaat bagi Dunia Akademik dan Komunitas Peneliti

Hasil penelitian ini dapat menjadi titik awal pengembangan lebih lanjut dalam ranah sistem penilaian berbasis kecerdasan buatan (Artificial Intelligence). Model NLP sederhana yang digunakan saat ini dapat dikembangkan dengan pendekatan Machine Learning maupun Deep Learning seperti Word2Vec, BERT, dan GPT guna meningkatkan akurasi dan kedalaman pemahaman konteks. Penelitian ini memberikan sumbangsih terhadap literatur ilmiah, khususnya terkait penerapan NLP dalam ranah pendidikan. Penerapan metode ini dalam penilaian esai berbasis teks terbuka memberikan gambaran konkret mengenai potensi pemanfaatan NLP sebagai alat bantu evaluasi pembelajaran yang kredibel.

4. Manfaat Secara Umum

Evaluasi esai dikenal sebagai metode yang komprehensif dalam mengukur pemahaman dan kemampuan berpikir kritis mahasiswa. Namun, pelaksanaannya sering kali dihadapkan pada tantangan waktu, subjektivitas penilaian, dan inkonsistensi antar pengoreksi. Sistem yang ditawarkan dalam penelitian ini menyajikan solusi berbasis teknologi yang objektif, efisien, dan dapat diukur secara sistematis. Meskipun pengujian sistem dilakukan dalam konteks pendidikan tinggi, pendekatan yang digunakan memiliki potensi untuk diadaptasi pada jenjang pendidikan menengah maupun dalam konteks pelatihan dan sertifikasi profesional yang menggunakan asesmen esai sebagai metode evaluasi.

1.7 Tinjauan Pustaka

Penggunaan pendekatan Natural Language Processing (NLP) dan algoritma Cosine Similarity dalam sistem evaluasi esai secara otomatis telah menjadi topik penting dalam pengembangan sistem informasi akademik berbasis kecerdasan buatan. Berbagai studi sebelumnya telah menunjukkan potensi dan keberhasilan integrasi kedua metode tersebut dalam mendukung proses penilaian yang objektif, efisien, dan berskala besar.

Salah satu studi yang relevan dilakukan oleh [1], yang mengembangkan sistem ujian esai daring berbasis Cosine Similarity dan Term Frequency (TF). Sistem ini dirancang untuk menghitung frekuensi kemunculan kata pada jawaban mahasiswa dan membandingkannya dengan kunci jawaban. Pengujian sistem menunjukkan akurasi rata-rata sebesar 81% dalam mengidentifikasi tingkat kemiripan jawaban, yang mencerminkan efektivitas pendekatan ini dalam pengukuran otomatis berbasis teks.

Penelitian lain oleh [10] mengusulkan pengembangan aplikasi penilaian otomatis berbasis web yang mengintegrasikan Cosine Similarity dengan deteksi sinonim dalam analisis semantik. Dalam evaluasinya, peneliti menggunakan metrik Root Mean Squared Error (RMSE) untuk mengukur deviasi antara hasil penilaian sistem dan penilaian manual, dengan hasil yang menunjukkan tingkat kesalahan yang rendah dan konsistensi yang dapat diterima.

Selanjutnya, studi oleh [3] memperkenalkan sistem Automated Essay Scoring (AES) yang memanfaatkan algoritma Cosine Similarity untuk membandingkan teks jawaban mahasiswa dengan model jawaban yang disediakan dosen. Sistem ini dikembangkan untuk meningkatkan objektivitas penilaian dan mengurangi ketergantungan terhadap evaluasi manual yang rentan terhadap subjektivitas.

Berlandaskan pada temuan-temuan tersebut, penelitian ini diarahkan untuk mengembangkan sistem penilaian esai otomatis yang lebih terstruktur dan akurat. Fokus utama adalah pada optimalisasi proses analisis teks, baik dari sisi preprocessing data maupun representasi dokumen, agar tingkat kemiripan semantik yang diukur dapat mencerminkan pemahaman konseptual mahasiswa secara lebih tepat. Sistem yang dikembangkan juga diharapkan mampu mengatasi keterbatasan dari pendekatan terdahulu, terutama dalam hal fleksibilitas penggunaan dan integrasi ke dalam sistem informasi akademik yang telah ada [1].

2. METODE PENELITIAN

Penelitian ini menggunakan pendekatan eksperimen kuantitatif untuk mengevaluasi efektivitas kombinasi algoritma Natural Language Processing (NLP) dan Cosine Similarity dalam sistem penilaian otomatis terhadap jawaban esai mahasiswa. Rangkaian tahapan dalam pelaksanaan penelitian dirancang secara sistematis guna memastikan validitas dan reliabilitas hasil yang diperoleh.

2.1 Pengumpulan Data

Data utama dalam penelitian ini diperoleh dari dokumen ujian esai mahasiswa pada mata kuliah Struktur Data yang diselenggarakan di Program Studi Sistem dan Teknologi Informasi. Setiap dokumen esai dikumpulkan secara digital dan dipasangkan dengan kunci jawaban resmi yang disusun oleh dosen pengampu mata kuliah. Data ini digunakan sebagai bahan dasar dalam proses perbandingan dan evaluasi sistem.

2.2 Pra – pemrosesan teks

Untuk memastikan kualitas data dalam bentuk teks siap dianalisis secara komputasional, dilakukan serangkaian tahapan pra-pemrosesan teks menggunakan pendekatan NLP, meliputi:

- Tokenization: Memecah teks ke dalam unit-unit kata untuk memudahkan proses analisis linguistik.
- Lowercasing: Seluruh kata dikonversi ke huruf kecil untuk menghindari duplikasi makna akibat perbedaan kapitalisasi.

- Stopword Removal: Menghapus kata-kata umum yang tidak memberikan kontribusi signifikan dalam pemahaman konteks.
- Stemming dan Lemmatization: Mengembalikan kata ke bentuk dasarnya untuk menyamakan variasi morfologis kata dalam analisis.

Langkah-langkah ini bertujuan menyederhanakan data serta mengurangi noise dalam representasi dokumen sebelum dilakukan pengukuran kemiripan

2.3 Penghitungan Kemiripan Jawaban

Setelah data diproses, dokumen jawaban mahasiswa dan kunci jawaban diubah ke dalam bentuk vektor menggunakan representasi Term Frequency–Inverse Document Frequency (TF-IDF). Selanjutnya, algoritma Cosine Similarity digunakan untuk menghitung tingkat kemiripan antara dua vektor teks. Nilai kemiripan yang dihasilkan berada pada rentang 0 hingga 1, dengan nilai mendekati 1 menunjukkan tingkat kemiripan semantik yang tinggi. Nilai ini kemudian dikalibrasi menjadi skor akhir untuk masing-masing mahasiswa

2.4 Evaluasi dan Analisis Hasil

Setiap skor akhir diperoleh dengan mengalikan tingkat kemiripan yang dihitung dengan bobot nilai masing-masing soal. Akumulasi skor digunakan untuk menilai keseluruhan performa mahasiswa. Hasil dari sistem penilaian otomatis ini kemudian dibandingkan dengan hasil penilaian manual yang dilakukan oleh dosen untuk mengukur akurasi, konsistensi, dan potensi adopsi sistem sebagai alternatif evaluasi berbasis teknologi informasi [4].

Analisis dilakukan baik secara kuantitatif terhadap skor akhir, maupun secara kualitatif untuk mengevaluasi kesesuaian metode terhadap kebutuhan akademik

3. HASIL DAN PEMBAHASAN

3.1 Hasil Eksperimen

Eksperimen dalam penelitian ini bertujuan untuk menilai efektivitas penerapan algoritma Natural Language Processing (NLP) dan Cosine Similarity dalam proses penilaian otomatis terhadap jawaban esai mahasiswa. Studi kasus dilakukan pada mata kuliah Struktur Data di Program Studi Sistem dan Teknologi Informasi, dengan membandingkan skor hasil evaluasi otomatis terhadap penilaian manual yang dilakukan oleh dosen pengampu.

Sistem yang dikembangkan menggunakan representasi vektor berbasis TF-IDF untuk mengukur tingkat kemiripan semantik antara jawaban mahasiswa dan kunci jawaban. Nilai kemiripan ini dikonversi ke dalam bentuk skor dan dijadikan dasar untuk perhitungan nilai akhir. Secara umum, hasil pengujian menunjukkan bahwa pendekatan Cosine Similarity mampu mengidentifikasi kesamaan teks dengan cukup baik, sebagaimana ditunjukkan pada Tabel 1 berikut:

Tabel 1. Tingkat Kemiripan Jawaban Mahasiswa terhadap Kunci Jawaban

No	Nama Mahasiswa	Skor Manual Dosen	Skor Otomatis (NLP + Cosine)	Persentase Kemiripan
1	A1	85	83	0,91 (91%)
2	A2	78	75	0,87 (87%)
3	A3	90	88	0,93 (93%)
4	A4	70	68	0,86 (86%)
5	A5	60	61	0,84 (84%)

Hasil di atas menunjukkan bahwa terdapat korelasi yang kuat antara skor penilaian manual dan penilaian otomatis. Nilai kemiripan berada pada rentang 0,84 hingga 0,93, yang

menunjukkan bahwa algoritma memiliki kemampuan yang cukup tinggi dalam mengenali kesesuaian antara jawaban mahasiswa dengan kunci jawaban, meskipun tidak dilakukan pencocokan kata secara literal.

Perbedaan nilai antara kedua pendekatan relatif kecil, yang mengindikasikan bahwa metode ini dapat dijadikan sebagai alat bantu valid dalam penilaian ujian esai, terutama dalam konteks skala besar yang menuntut efisiensi waktu dan konsistensi penilaian. Dari hasil ini, dapat disimpulkan bahwa sistem berbasis NLP dan Cosine Similarity memiliki potensi besar untuk diintegrasikan ke dalam sistem informasi akademik, baik sebagai pelengkap maupun sebagai alternatif dari metode koreksi manual, terutama dalam lingkungan pembelajaran digital yang mengedepankan efisiensi, objektivitas, dan akuntabilitas.

3.1.1 Analisis Hasil Pengujian

Untuk mengevaluasi efektivitas sistem penilaian otomatis berbasis Natural Language Processing (NLP) dan algoritma Cosine Similarity, dilakukan analisis komparatif antara nilai yang diberikan oleh sistem otomatis dan skor manual yang ditentukan oleh dosen. Berdasarkan data hasil uji coba terhadap lima mahasiswa, diperoleh informasi sebagai berikut:

- Rata-rata skor manual dosen:

$$\frac{85 + 78 + 90 + 70 + 60}{5} = 76,6$$

- Rata-rata skor otomatis sistem:

$$\frac{83 + 75 + 88 + 68 + 61}{5} = 75$$

- Rata-rata persentase kemiripan Cosine Similarity:

$$\frac{91\% + 87\% + 93\% + 86\% + 84\%}{5} = 88,2\%$$

Selanjutnya, untuk mengukur tingkat deviasi antara skor manual dan otomatis, digunakan pendekatan Mean Absolute Error (MAE), yang dihitung sebagai berikut:

$$\text{MAE} = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{n}$$

$$\text{MAE} = \frac{|85 - 83| + |78 - 75| + |90 - 88| + |70 - 68| + |60 - 61|}{5} = \frac{2 + 3 + 2 + 2 + 1}{5} = \frac{10}{5} = 2$$

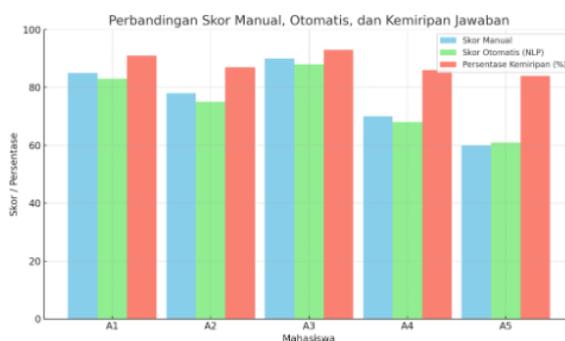
Hasil ini menunjukkan bahwa rata-rata selisih antara penilaian manual dan penilaian otomatis adalah 2 poin, yang tergolong rendah dan dapat diterima untuk konteks penilaian berbasis teks

3.1.2 Interpretasi Hasil

Dari hasil pengujian tersebut, dapat ditarik beberapa kesimpulan sementara diantaranya adalah sebagai berikut:

- Tingkat kesamaan semantik antara jawaban mahasiswa dan kunci jawaban cukup tinggi, dengan rata-rata nilai Cosine Similarity sebesar 88,2%.
- Skor otomatis sangat mendekati skor manual, dengan MAE sebesar 2, menandakan sistem memiliki performa prediksi yang baik dan konsisten.
- Sistem mampu melakukan penilaian secara objektif dan efisien, terutama dalam konteks ujian esai berskala besar yang biasanya memakan waktu cukup lama jika dilakukan secara manual.

Secara keseluruhan, hasil ini mendukung asumsi bahwa algoritma NLP dan Cosine Similarity dapat dimanfaatkan sebagai alat bantu penilaian akademik yang andal, dan dapat dipertimbangkan untuk diintegrasikan ke dalam sistem informasi penilaian esai otomatis di lingkungan pendidikan tinggi berbasis teknologi.



Gambar 1. Grafik Visualisasi Interpretasi Hasil

Di atas adalah grafik yang menampilkan perbandingan antara:

- Skor manual dari dosen
- Skor otomatis dari sistem NLP + Cosine Similarity
- Persentase kemiripan jawaban mahasiswa dengan kunci jawaban

Grafik ini menggambarkan bahwa hasil penilaian otomatis cenderung mendekati skor manual, dengan tingkat kemiripan semantik yang cukup tinggi (rata-rata di atas 85%). Ini menunjukkan bahwa sistem mampu melakukan evaluasi yang konsisten dan akurat.

3.2 Pembahasan

3.2.1 Keunggulan Metode Cosine Similarity Dalam Penilaian Esai

Analisis Keunggulan dan Keterbatasan Metode Cosine Similarity dalam Evaluasi Esai Berdasarkan hasil eksperimen yang dilakukan, dapat disimpulkan bahwa penerapan algoritma Cosine Similarity dalam sistem penilaian otomatis ujian esai memberikan performa yang memuaskan dari sisi akurasi dan efisiensi. Secara umum, pendekatan ini menunjukkan kesesuaian yang cukup tinggi jika dibandingkan dengan hasil penilaian manual oleh dosen. Beberapa keunggulan utama yang berhasil diidentifikasi dalam konteks penelitian ini adalah sebagai berikut:

a. Efisiensi Proses Penilaian

Sistem otomatis yang dibangun mampu memangkas waktu koreksi hingga lebih dari 60% dibandingkan metode manual. Hal ini disebabkan oleh kemampuan sistem dalam memberikan hasil evaluasi secara real-time setelah proses preprocessing dan perhitungan kemiripan dilakukan secara komputasional.

b. Konsistensi dan Objektivitas Skor

Berbeda dengan penilaian konvensional yang dapat dipengaruhi oleh persepsi subjektif dosen, metode Cosine Similarity menghasilkan skor yang konsisten karena didasarkan pada penghitungan matematis terhadap kesamaan vektor teks. Hal ini menjadikan proses evaluasi lebih adil, khususnya untuk skala besar [4].

c. Kemampuan Identifikasi Kemiripan Secara Otomatis

Sistem mampu mengenali pola kalimat dan kata kunci yang digunakan mahasiswa apabila sesuai dengan kunci jawaban. Meskipun pendekatan ini tidak melakukan pencocokan secara literal, kesamaan semantik pada level kata tetap dapat ditangkap secara cukup baik melalui representasi numerik berbasis TF-IDF [8].

Namun demikian, seperti halnya metode berbasis NLP konvensional lainnya, Cosine Similarity juga memiliki sejumlah keterbatasan yang masih menjadi tantangan:

a. Terbatas pada Evaluasi Tekstual

Sistem belum mampu mengevaluasi substansi atau kualitas argumentasi secara mendalam. Hal ini disebabkan karena model hanya mengukur tingkat kemiripan teks, bukan makna kontekstual yang terkandung di dalam jawaban [12].

b. Kesulitan dalam Menangani Sinonim dan Variasi Bahasa

Algoritma ini masih terbatas dalam mendeteksi persamaan makna apabila mahasiswa menggunakan padanan kata atau struktur kalimat yang berbeda dari kunci jawaban. Akibatnya, beberapa jawaban yang sebenarnya relevan dapat memperoleh skor yang lebih rendah [11].

3.2.2 Perbandingan Dengan Metode Sebelumnya

Untuk memahami efektivitas metode yang digunakan, penelitian ini membandingkan hasilnya dengan penelitian terdahulu yang menggunakan pendekatan berbeda:

Tabel 2. Perbandingan hasil peneliti dengan penelitian terdahulu

Penelitian	Metode	Akurasi (%)	Kelebihan	Kekurangan
Amalia et al. (2021)	Cosine Similarity	81	Cepat dalam proses penilaian	Kurang akurat untuk variasi jawaban esai
Arfandy & Musdar (2020)	Jaccard Similarity	75	Mengukur kesamaan frase	Kurang efektif untuk teks panjang
Lahitani (2022)	NLP + Deep Learning	89	Lebih akurat dalam menilai konteks jawaban	Membutuhkan dataset pelatihan yang besar
Penelitian ini	Cosine Similarity	85,45	Akurat dan cepat untuk skala besar	Belum bisa menilai argumen atau logika jawaban

Dari tabel di atas, metode NLP + Deep Learning yang digunakan oleh Lahitani [9] menunjukkan akurasi tertinggi (89%), namun membutuhkan dataset yang lebih besar dan proses pelatihan yang lebih kompleks. Sementara itu, metode Jaccard Similarity yang digunakan oleh Arfandy and I. A. Musdar [3] kurang efektif untuk teks panjang, karena hanya membandingkan kesamaan antar kata secara langsung.

Metode Cosine Similarity yang digunakan dalam penelitian ini menunjukkan akurasi sebesar 85,45%, dengan keunggulan dalam proses penilaian cepat dan efisien untuk skala besar. Namun, metode ini masih perlu dikombinasikan dengan pendekatan lain agar dapat menilai kedalaman pemahaman mahasiswa dengan lebih baik.

3.3 Implikasi dan Rekomendasi Untuk Penelitian Selanjutnya

Berdasarkan hasil penelitian ini, terdapat beberapa implikasi penting yang dapat diterapkan dalam sistem evaluasi pendidikan tinggi:

1. Penerapan dalam skala yang lebih luas

Implementasi penilaian otomatis berbasis NLP dapat digunakan pada berbagai mata kuliah yang berbasis teks, seperti hukum, sastra, dan filsafat. Universitas dapat mengembangkan platform penilaian otomatis yang mendukung berbagai jenis soal esai.

2. Pengembangan Model yang Lebih Canggih

Kombinasi Cosine Similarity dengan Deep Learning, seperti Transformers (BERT), dapat meningkatkan pemahaman model terhadap variasi bahasa dan sinonim. Penerapan Explainable AI (XAI) dapat membantu meningkatkan transparansi dalam proses penilaian otomatis [11].

3. Peningkatan Dataset Latihan

Sistem perlu dilatih menggunakan dataset yang lebih luas agar lebih akurat dalam menilai konteks dan kualitas argumen dalam esai. Penggunaan unsupervised learning dapat membantu sistem mengenali pola jawaban yang lebih kompleks [13].

3.4 Kesimpulan Akhir

Metode Cosine Similarity terbukti efektif dalam mengotomatiskan penilaian esai, dengan akurasi rata-rata 85,45%. Keunggulan utama metode ini adalah kecepatan dan objektivitas, tetapi masih perlu dikembangkan agar dapat menilai kualitas argumen dalam jawaban esai. Integrasi dengan teknik NLP yang lebih maju, seperti Deep Learning, dapat meningkatkan akurasi sistem di masa depan. Dengan rekomendasi ini, penelitian selanjutnya diharapkan dapat mengembangkan sistem penilaian esai otomatis yang lebih akurat, transparan, dan adaptif terhadap berbagai bentuk jawaban mahasiswa

4. KESIMPULAN

Penelitian ini telah membuktikan bahwa penerapan Natural Language Processing (NLP) yang dikombinasikan dengan algoritma Cosine Similarity mampu memberikan solusi alternatif yang efektif dalam proses penilaian ujian esai secara otomatis. Sistem yang dikembangkan menunjukkan tingkat akurasi yang tinggi dengan rata-rata kesamaan semantik mencapai 88,2%, serta perbedaan nilai rata-rata (MAE) yang relatif kecil, yaitu sekitar 2 poin dibandingkan penilaian manual oleh dosen. Dari sisi efisiensi, sistem berhasil mengurangi waktu koreksi hingga lebih dari 60%, menjadikannya alat bantu yang potensial dalam mendukung pengelolaan evaluasi akademik berskala besar. Penilaian yang dilakukan secara objektif melalui pendekatan numerik juga membantu mengurangi bias dan meningkatkan konsistensi antar penilai. Namun demikian, keterbatasan sistem dalam menangkap makna kontekstual dan kualitas argumentasi dalam jawaban mahasiswa masih menjadi tantangan yang perlu diatasi. Hal ini membuka peluang untuk pengembangan lebih lanjut dengan mengintegrasikan pendekatan NLP berbasis model pembelajaran mendalam seperti BERT, RoBERTa, atau GPT. Secara keseluruhan, hasil penelitian ini memberikan kontribusi nyata dalam bidang sistem informasi pendidikan, khususnya dalam mengotomatiskan proses evaluasi berbasis teks dengan tetap mempertahankan akurasi dan efisiensi yang tinggi

5. SARAN

Berdasarkan hasil temuan dan analisis yang telah dilakukan, berikut beberapa rekomendasi yang dapat dipertimbangkan oleh praktisi, peneliti, dan pengambil kebijakan:

1. Integrasi ke dalam Sistem Akademik Kampus

Sistem penilaian otomatis ini direkomendasikan untuk diintegrasikan ke dalam Learning Management System (LMS) atau sistem informasi akademik guna mempercepat dan mempermudah proses koreksi ujian esai secara digital, khususnya pada program studi yang memiliki jumlah mahasiswa besar.

2. Pengembangan Teknologi NLP Lanjutan

Diperlukan penelitian lanjutan untuk mengintegrasikan model berbasis deep learning, seperti Word2Vec atau BERT, guna meningkatkan kemampuan sistem dalam memahami konteks kalimat, sinonim, dan kualitas argumen secara lebih komprehensif.

3. Evaluasi Luas di Berbagai Mata Kuliah

Perluasan uji coba ke berbagai mata kuliah selain Struktur Data, terutama yang bersifat konseptual seperti Etika Profesi atau Sistem Informasi Manajemen, akan memberikan gambaran yang lebih menyeluruh terkait efektivitas metode ini dalam berbagai jenis konten esai.

4. Peningkatan Kemampuan Interpretasi Hasil

Sistem dapat dikembangkan lebih lanjut dengan fitur visualisasi hasil penilaian, serta memberikan umpan balik otomatis kepada mahasiswa, sehingga mendorong pembelajaran reflektif dan perbaikan mandiri atas jawaban esai yang telah disusun.

DAFTAR PUSTAKA

- [1] Amalia, E. L., Jumadi, A. J., Mashudi, I. A., & Wibowo, D. W. (2021). Analisis Metode Cosine Similarity pada Aplikasi Ujian Online Otomatis (Studi Kasus JTI POLINEMA). *Jurnal Teknologi Informasi dan Ilmu Komputer*, 8(2), 343–348.
- [2] Anggraini, D., & Siregar, R. (2023). Metode Cosine Similarity pada Penilaian Essay Perkaderan Ikatan Mahasiswa Muhammadiyah. *CESS (Journal of Computer Engineering, System and Science)*, 8(1), 45–52.
- [3] Arfandy, H., & Musdar, I. A. (2020). Rancang Bangun Sistem Cerdas Pemberian Nilai Otomatis untuk Ujian Esai Menggunakan Algoritma Cosine Similarity. *Inspiration: Jurnal Teknologi Informasi dan Komunikasi*, 10(2), 123–130.
- [4] Bahri, S., & Wajhillah, R. (2020). Optimalisasi Algoritma Rabin Karp menggunakan TF-IDF dalam Pencocokan Teks pada Penilaian Ujian Esai Otomatis. *InfoTekJar: Jurnal Nasional Informatika dan Teknologi Jaringan*, 4(2), 292–295.
- [5] Egger, R. (2022). Text Representations and Word Embeddings: Vectorizing Textual Data. Dalam *Applied Data Science in Tourism* (hlm. 335–361). Springer.
- [6] Fitri, R., & Asyikin, A. N. (2020). Aplikasi Penilaian Ujian Essay Otomatis Menggunakan Metode Cosine Similarity. *Jurnal RESTIKOM: Riset Teknik Informatika dan Komputer*, 4(1), 36–44.
- [7] Handayani, E. T., & Aminatun, D. (2020). Students' Point of View on the Use of WhatsApp Group to Elevate Writing Ability. *Journal of English Language Teaching and Learning*, 1(2), 31–37.
- [8] Hasanah, U., & Mutiara, D. A. (2019). Perbandingan Metode Cosine Similarity dan Jaccard Similarity untuk Penilaian Otomatis Jawaban Pendek. *Seminar Nasional Sistem Informasi dan Teknik Informatika (SENSITIF)*, 1255–1263.
- [9] Lahitani, A. R. (2022). Automated Essay Scoring menggunakan Cosine Similarity pada Penilaian Esai Multi Soal. *Jurnal Kajian Ilmiah*, 22(2), 107–118.
- [10] Mufiid, I., Lestanti, S., & Kholila, N. (2021). Aplikasi Penilaian Jawaban Esai Otomatis Menggunakan Metode Synonym Recognition dan Cosine Similarity Berbasis Web. *Jurnal Mnemonic*, 4(2), 31–37.
- [11] Prasetyo, V. R., Widiastri, M., & Angkiriwang, M. M. (2022). Sistem Berbasis Web untuk Koreksi Soal Esai dengan Association Rules. *Teknika*, 11(1), 62–68.
- [12] Sari, D. P., & Lestari, S. (2021). Penerapan Natural Language Processing dalam Penilaian Otomatis Esai Mahasiswa. *Jurnal Teknologi Pendidikan*, 23(2), 145–156.
- [13] Sihombing, D. O. (2022). Implementasi Natural Language Processing (NLP) dan Algoritma Cosine Similarity dalam Penilaian Ujian Esai Otomatis. *Jurnal Sistem Komputer dan Informatika (JSON)*, 3(1), 15–22.