

PEMANFAATAN ALGORITMA *COSINE SIMILARITY* UNTUK MENGGKOREKSI UJIAN ESAI

Reno Rana Raffi¹⁾, Risky Aswi Ramadhani^{*2)}, Ardi Sanjaya³⁾

1. Teknik Informatika, Fakultas Teknik dan Ilmu Komputer, Universitas Nusantara PGRI Kediri

2. Teknik Informatika, Fakultas Teknik dan Ilmu Komputer, Universitas Nusantara PGRI Kediri

3. Teknik Informatika, Fakultas Teknik dan Ilmu Komputer, Universitas Nusantara PGRI Kediri

Article Info

Kata Kunci: penilaian esai otomatis; pemrosesan bahasa alami; *cosine similarity*.

Keywords: *automated essay scoring; natural language processing; cosine similarity*.

Article history:

Received 26 May 2025

Revised 4 June 2025

Accepted 5 July 2025

Available online 1 September 2025

DOI :

<https://doi.org/10.29100/jipi.v10i3.8058>

* Corresponding author.

Corresponding Author

E-mail address:

riskyaswiramadhani@gmail.com

ABSTRAK

Penilaian ujian esai merupakan tantangan dalam dunia pendidikan karena adanya variasi bentuk jawaban yang memerlukan pemeriksaan secara menyeluruh oleh guru. Peningkatan jumlah peserta didik juga menambah beban kerja dan memperlambat proses koreksi dalam penilaian ujian. Penelitian ini bertujuan untuk mengembangkan sistem penilaian otomatis menggunakan metode *Term Frequency Inverse Document Frequency* dan algoritma *Cosine Similarity*, yang diimplementasikan di SMK Intensif Baitussalam Tanjunganom. Data berupa jawaban siswa dan kunci jawaban guru diproses melalui tahapan *preprocessing*, meliputi *cleaning, case folding, tokenization, stopword*, dan *stemming*. Sistem menghitung tingkat kemiripan jawaban menggunakan *Term Frequency Inverse Document Frequency* dan *Cosine Similarity* untuk menentukan skor akhir. Evaluasi dilakukan pada soal esai mata pelajaran sejarah dengan pendekatan klasifikasi biner menggunakan *threshold*. Hasil dari pengujian menunjukkan bahwa sistem menghasilkan akurasi sebesar 80%, presisi 100%, recall 80%, dan F1-score sebesar 89%. Temuan ini membuktikan bahwa sistem mampu memberikan penilaian yang objektif dan efisien, serta berpotensi menjadi solusi dalam digitalisasi proses evaluasi pembelajaran secara otomatis.

ABSTRACT

Essay assessment presents a challenge in the field of education due to the wide variety of student responses, which require thorough examination by teachers. The increasing number of students also adds to the workload and slows down the grading process. This study aims to develop an automated essay scoring system using the Term Frequency Inverse Document Frequency method and Cosine Similarity algorithm, implemented at SMK Intensif Baitussalam Tanjunganom. The data, consisting of student answers and teacher answer keys, were processed through several preprocessing stages, including cleaning, case folding, tokenization, stopword removal, and stemming. The system calculates the similarity between student answers and the answer key using Term Frequency Inverse Document Frequency and Cosine Similarity to determine the final score. Evaluation was conducted on essay questions from a history subject using a binary classification approach with a defined threshold. The results showed that the system achieved an accuracy of 80%, precision of 100%, recall of 80%, and an F1-score of 89%. These findings demonstrate that the system can provide objective and efficient assessments and has the potential to support the digitalization of the learning evaluation process automatically.

I. PENDAHULUAN

Perkembangan teknologi informasi dan komunikasi pada era digital telah membawa perubahan signifikan di berbagai sektor, termasuk bidang pendidikan. Integrasi teknologi dalam pendidikan tidak lagi menjadi sekadar tren, melainkan sebuah kebutuhan untuk meningkatkan efisiensi dan efektivitas dalam proses pembelajaran, evaluasi, serta manajemen akademik. Salah satu aspek penting dalam dunia pendidikan adalah proses evaluasi pembelajaran, di mana ujian digunakan untuk mengukur tingkat pemahaman dan kompetensi siswa terhadap materi yang telah diajarkan [1]. Salah satu bentuk evaluasi yang umum digunakan adalah ujian esai, yang memungkinkan siswa untuk menjawab pertanyaan dengan penjabaran berdasarkan pemahaman dan penalaran

mereka. Bentuk soal ini memiliki keunggulan dalam menggali kemampuan berpikir kritis siswa, karena memungkinkan mereka menyampaikan ide, argumen, dan pemahamannya secara bebas [2].

Di tingkat sekolah menengah, seperti SMK Intensif Baitussalam Tanjunganom, ujian esai masih menjadi salah satu instrumen utama dalam Ujian Tengah Semester (UTS) dan Ujian Akhir Semester (UAS). Urgensi dari permasalahan ini menjadi lebih nyata ketika jumlah siswa yang mengikuti ujian semakin banyak, sehingga guru membutuhkan waktu lebih lama untuk melakukan koreksi satu per satu secara manual [3]. Selain memperlambat penyampaian hasil belajar kepada siswa, proses ini juga berisiko menurunkan konsistensi penilaian akibat kelelahan atau persepsi individu yang berbeda [4]. Permasalahan ini tidak hanya relevan di lingkungan SMK, tetapi juga memiliki urgensi yang luas pada berbagai tingkatan pendidikan, termasuk sekolah dasar, menengah, hingga pendidikan tinggi. Dengan meningkatnya pembelajaran daring dan penilaian berbasis digital, teknologi *Natural Language Processing (NLP)* memiliki potensi untuk diterapkan sebagai solusi penilaian otomatis yang efisien dan objektif di berbagai platform evaluasi.

Dalam konteks inilah, teknologi *Natural Language Processing (NLP)* dapat menjadi solusi yang relevan dan adaptif untuk berbagai tingkatan pendidikan. *Natural Language Processing (NLP)* merupakan cabang dari kecerdasan buatan yang memungkinkan komputer untuk memahami, memproses, dan menganalisis bahasa manusia secara alami. Dalam konteks penilaian esai, *Natural Language Processing (NLP)* dapat digunakan untuk menilai kemiripan jawaban siswa dengan kunci jawaban secara objektif dan otomatis, sehingga dapat mengurangi beban guru sekaligus meningkatkan kecepatan dan konsistensi evaluasi.

Salah satu pendekatan yang umum digunakan dalam *Natural Language Processing (NLP)* adalah algoritma *Cosine Similarity*, yang menghitung kemiripan antara dua dokumen teks dalam bentuk representasi vektor. Metode ini bekerja dengan membandingkan sudut kosinus antara dua vektor yang telah dibobot menggunakan metode *Term Frequency (TF)*. Keunggulan *Cosine Similarity* terletak pada kemampuannya untuk mengukur kemiripan konten meskipun teks memiliki panjang yang berbeda. Dibandingkan dengan metode lain seperti *Jaccard Similarity* atau *Euclidean Distance*, *Cosine Similarity* lebih efektif dalam mempertahankan konteks semantik pada teks [5].

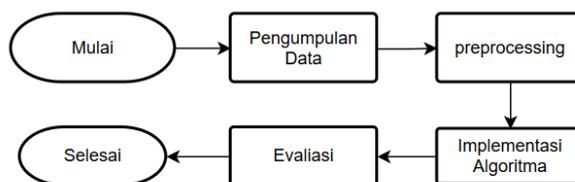
Meskipun beberapa penelitian sebelumnya telah menerapkan metode ini, masih terdapat keterbatasan dalam menangani kompleksitas bahasa, seperti penggunaan sinonim, variasi struktur kalimat, dan gaya bahasa siswa. Misalnya, pada penelitian terdahulu mengenai aplikasi *Cerdas Cermat* yang menggunakan *Cosine Similarity* dan database sinonim, sistem belum mampu menangani konteks penggunaan kata dalam kalimat yang lebih kompleks secara optimal [6]. Penelitian lainnya menggunakan pendekatan *syntactic-semantic similarity* dalam penilaian esai Bahasa Indonesia dan memperoleh akurasi sekitar 70,44% terhadap penilaian manual [7]. Namun demikian, penelitian terdahulu masih memiliki kelemahan dalam memahami kompleksitas bahasa alami, seperti variasi struktur kalimat, penggunaan sinonim dalam konteks tertentu, serta gaya bahasa individual siswa yang dapat mempengaruhi makna keseluruhan jawaban. Hal ini menunjukkan bahwa pendekatan yang digunakan masih perlu disempurnakan untuk menghasilkan penilaian yang benar-benar mencerminkan pemahaman siswa secara semantik.

Berdasarkan identifikasi terhadap keterbatasan penelitian sebelumnya, penelitian ini bertujuan untuk mengembangkan sistem penilaian otomatis esai menggunakan *Natural Language Processing (NLP)* dengan menerapkan algoritma *Cosine Similarity* yang dipadukan dengan tahapan *text preprocessing* yang lebih komprehensif. Tahapan tersebut meliputi *tokenisasi*, *normalisasi*, penghilangan *stopwords*, *stemming*, dan pembobotan kata menggunakan metode *Term Frequency*. Kemudian, sistem akan menghitung tingkat kemiripan antara jawaban siswa dan kunci jawaban berdasarkan nilai *Cosine Similarity*.

Untuk mengatasi permasalahan tersebut, pemanfaatan teknologi berbasis *Natural Language Processing (NLP)* menjadi solusi yang menjanjikan. *Natural Language Processing (NLP)* telah digunakan secara luas di berbagai bidang, seperti pada sistem *chatbot* untuk layanan pelanggan, analisis sentimen di media sosial, serta sistem pencarian berbasis teks di mesin pencari. Keberhasilan penerapan *Natural Language Processing (NLP)* dalam memahami dan memproses bahasa alami di bidang-bidang tersebut menunjukkan potensi besar teknologi ini untuk diadaptasi dalam dunia pendidikan, khususnya dalam penilaian esai otomatis. Melalui pendekatan ini, diharapkan sistem dapat memberikan alternatif solusi dalam meningkatkan efisiensi dan objektivitas penilaian esai, serta memberikan kontribusi nyata dalam mendukung transformasi digital di dunia pendidikan, khususnya dalam hal evaluasi pembelajaran. Sistem ini juga diharapkan dapat diterapkan di berbagai jenjang pendidikan lain sebagai inovasi penunjang tugas guru dan peningkatan mutu pembelajaran.

II. METODE PENELITIAN

Dalam tahapan penelitian, peneliti menggunakan beberapa tahapan-tahapan. Berikut alur penelitian yang akan dilakukan seperti pada gambar 1.



Gambar 1 Tahapan Penelitian

A. Pengumpulan data

Berdasarkan metode penelitian yang digunakan, tahapan pertama yang dilakukan adalah proses pengumpulan data yang bertujuan untuk memahami kebutuhan apa saja yang diperlukan untuk mendukung penelitian. Tahap ini mencakup pengumpulan materi mata pelajaran sejarah di SMK intensif baitussalam. Data yang akan dikumpulkan meliputi soal ujian dari guru, kunci jawaban yang dibuat guru, serta jawaban dari siswa SMK intensif baitussalam yang mengikuti ujian. Tahap berikutnya adalah pengembangan aplikasi, yang dirancang untuk menyelenggarakan ujian esai untuk menilai jawaban siswa menggunakan algoritma *Cosine Similarity*. Pendekatan ini dipilih karena mampu memberikan hasil yang lebih objektif dan sistematis dalam menilai kualitas isi teks berdasarkan kesamaan makna dan struktur kata.

B. Text Preprocessing

Tahapan awal dalam *Natural Language Processing* adalah *Text preprocessing* merupakan proses mempersiapkan dan membersihkan teks mentah sehingga siap untuk digunakan dalam model atau analisis *Natural Language Processing (NLP)*. Tujuan dari *preprocessing* adalah mengubah teks ke bentuk yang lebih terstruktur dan standar, sehingga algoritma dapat bekerja lebih efisien dan akurat[8]. Adapun beberapa langkah untuk melakukan tahapan *text Preprocessing* adalah sebagai berikut :

1. *Cleaning* Proses mengkonversi teks ke format standar dengan menghilangkan karakter khusus seperti simbol "@", "%", "\$", "!", "?" dan sejenisnya, akan dihapus guna meningkatkan keakuratan hasil pengolahan data [9].
2. *Case folding* Merupakan proses normalisasi dengan mengubah semua huruf dalam teks menjadi huruf kecil *lowercase* agar tidak terjadi perbedaan pengenalan kata hanya karena perbedaan kapitalisasi. Misalnya, kata "Indonesia" dan "indonesia" akan dianggap sama [10].
3. *Tokenization* adalah proses memisahkan teks menjadi unit-unit yang lebih kecil, seperti kata-kata, frasa, atau token. Proses tersebut merupakan proses dalam pemrosesan teks. Misalnya, kalimat "Saya sedang mempelajari teori pemrograman" akan di pisahkan menjadi seperti ("saya", "sedang", "mempelajari", "teori", "pemrograman") [11].
4. *Stopword Removal* adalah teknik untuk menghilangkan kata-kata yang dianggap tidak penting, seperti kata sambung, kata depan, kata ganti, atau kata-kata yang tidak relevan untuk analisis sentimen, sehingga perhitungan dapat berkonsentrasi pada kata-kata yang lebih relevan. Tahap ini menghapus kata-kata umum yang tidak memiliki makna penting dalam analisis, seperti "dan", "yang", "di", "ke", dan sebagainya. *Stopword* dihapus agar algoritma lebih fokus pada kata-kata kunci yang bermakna [12].
5. *Stemming* menghapus imbuhan kata untuk menghasilkan bentuk kata dasar, baik berupa awalan ataupun akhiran. Proses ini mengubah kata berimbuhan menjadi bentuk dasarnya. Misalnya, kata "berlari", "lari-lari", dan "pelari" akan diubah menjadi "lari". Proses ini menggunakan daftar kata dasar *basicword list* [13].

C. Implementasi Algoritma

Setelah melakukan *preprocessing* terhadap kunci jawaban dan jawaban siswa tahapan selanjutnya adalah perhitungan menggunakan *term frequency inverse document frequency*. Selain itu, menggunakan algoritma *Cosine Similarity* untuk mengetahui seberapa mirip jawaban peserta ujian dengan kunci jawaban.

1) Term Frequency Inverse Document Frequency

Term Frequency (TF) adalah salah satu komponen utama yang berfungsi untuk mengukur seberapa sering suatu kata muncul dalam sebuah dokumen. Secara sederhana, *term frequency* menghitung frekuensi kemunculan suatu kata dalam dokumen tertentu dan memberikan gambaran tentang pentingnya kata tersebut dalam konteks dokumen tersebut. Semakin sering sebuah kata muncul dalam dokumen, semakin besar nilai *TF* yang diberikan pada kata tersebut [14] [15] [16].

$$WTF(ti, dj) = f(ti, dj) \quad (1)$$

$WTF(t_i, d_j)$ adalah nilai tf dari $term$ ke i pada dokumen ke j , sedangkan pada (t_i, d_j) adalah jumlah kemunculan dari $term$ ke i pada dokumen j .

Proses selanjutnya yaitu *Inverse Document Frequency (IDF)* bertujuan untuk mengukur tingkat keunikan suatu $term$ dalam kumpulan dokumen. Jika sebuah kata muncul di banyak dokumen, maka nilai IDF-nya akan rendah, menandakan bahwa kata tersebut kurang informatif. Sebaliknya, jika sebuah kata hanya muncul di sedikit dokumen, maka nilai IDF-nya tinggi, menunjukkan bahwa kata tersebut memiliki nilai diskriminatif yang lebih besar. Rumus perhitungan IDF adalah sebagai berikut:

$$IDF(t_i) = \log \left(\frac{N}{df(t_i)} \right) \quad (2)$$

Keterangan:

- $IDF(t_i)$ adalah nilai IDF dari $term$ $t(t_i)$
- N adalah total jumlah dokumen dalam koleksi,
- $df(t_i)$ adalah jumlah dokumen yang mengandung $term$ $t(t_i)$.

Gabungan dari kedua metode ini dikenal dengan nama TF-IDF (Term Frequency – Inverse Document Frequency). TF-IDF digunakan untuk menilai pentingnya suatu kata dalam satu dokumen relatif terhadap seluruh koleksi dokumen. Nilai TF-IDF dihitung dengan mengalikan nilai TF dengan nilai IDF dari suatu $term$, yaitu:

$$TF-IDF(t_i, d_j) = TF(t_i, d_j) \times IDF(t_i) \quad (3)$$

Dengan pendekatan ini, $TF-IDF$ mampu memberikan bobot yang lebih tinggi pada kata-kata yang sering muncul dalam suatu dokumen tetapi jarang muncul di dokumen lain, sehingga dapat digunakan secara efektif.

2) Algoritma Cosine similarity

Pemilihan *algoritma Cosine Similarity* dalam penelitian ini didasarkan pada kemampuannya yang sederhana namun efektif dalam menghitung kemiripan antar dokumen berbasis *representasi vektor*. Dibandingkan metode lain seperti *Jaccard Similarity* yang hanya memperhitungkan kesamaan elemen, *Cosine Similarity* memperhatikan arah vektor dari dua dokumen tanpa dipengaruhi oleh panjang dokumen [17]. Hal ini penting karena dalam esai, panjang jawaban siswa dapat bervariasi, namun makna yang dikandung bisa tetap serupa. Dengan membandingkan sudut antar vektor, *Cosine Similarity* mampu menangani variasi dalam ekspresi bahasa selama kata-kata kunci tetap konsisten [18].

Cosine Similarity adalah teknik yang digunakan untuk mengukur kesamaan antara dua vektor teks dalam ruang vektor. *Cosine Similarity* mengukur sudut antara dua vektor teks dan menghasilkan nilai antara 0 dan 1, di mana nilai 1 menunjukkan kesamaan sempurna antara dua teks, dan nilai 0 menunjukkan tidak ada kesamaan [19][20].

$$\text{sim} = \cos \theta \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n A_i^2 \times \sum_{i=1}^n B_i^2}} \quad (4)$$

Keterangan :

A = vektor

B = vektor

A_i = bobot $term$ i dalam blok A_i

B_i = bobot $term$ i dalam blok B_i

i = jumlah $term$ dalam kalimat

n = jumlah vektor

Dimana A ialah bobot setiap ciri pada vector A dan B merupakan bobot setiap ciri pada vector B . Apabila dikaitkan dengan *information retrieval* maka B adalah bobot pada setiap istilah dokumen B sedangkan A merupakan bobot pada setiap istilah dokumen A . Nilai kemiripan teks dapat dihitung dengan membandingkan dokumen 2 dengan dokumen 1. Nilai $A_i \cdot B_i$ adalah nilai yang diperoleh dari penjumlahan kata A dan B nilai A_i^2 diperoleh dari semua nilai kata dalam dokumen A , yang semua nilainya dipangkatkan dua begitu juga dengan kata B_i^2 , semua nilai yang diperoleh dipangkatkan dua kemudian dijumlahkan.

D. Evaluasi

Evaluasi sistem bertujuan untuk mengukur kinerja model dalam melakukan penilaian jawaban esai dibandingkan dengan penilaian manual oleh guru. Salah satu metode yang digunakan untuk menilai performa klasifikasi adalah *Confusion Matrix*, yang berfungsi untuk mengukur tingkat akurasi dan ketepatan klasifikasi sistem. Untuk menghitung hasil evaluasi terdapat 4 atribut [21][22], yaitu:

TABEL I
 CONFUSION MATRIX

Actual value	Predicted value	
	Positive (1)	Negative(0)
Positive(1)	TP	FP
Negative(0)	FN	TN

1. *True Positive (TP)*: Jumlah jawaban yang benar dinilai benar oleh sistem.
2. *False Positive (FP)*: Jawaban yang salah namun dinilai benar oleh sistem.
3. *True Negative (TN)*: Jawaban yang salah dinilai salah oleh sistem.
4. *False Negative (FN)*: Jawaban yang sebenarnya benar namun dinilai salah oleh sistem.

Keakuratan evaluasi dihitung melalui beberapa metrik, yaitu *recall*, *precision*, *accuracy* dan *F1-Score*.

1. *Accuracy* menunjukkan persentase keseluruhan prediksi yang benar. Berikut rumus perhitungan *accuracy*:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (5)$$

2. *Precision* mengukur seberapa akurat sistem dalam mengklasifikasikan jawaban benar. Berikut rumus perhitungan *Precision*:

$$Precision = \frac{TP}{TP+FP} \quad (6)$$

3. *Recall* mengukur kemampuan sistem dalam menemukan semua jawaban benar. Berikut rumus perhitungan *Recall*:

$$Recall = \frac{TP}{TP+FN} \quad (7)$$

4. *F1-Score* adalah metrik evaluasi yang dihitung dengan cara mengalikan nilai *precision* dan *recall*, kemudian membaginya dengan hasil penjumlahan antara *precision* dan *recall*. Perhitungan ini bertujuan untuk memperoleh keseimbangan optimal.

$$F1-Score = 2 \frac{(precision \times recall)}{(precision+recall)} \quad (8)$$

III. HASIL DAN PEMBAHASAN

A. Pengumpulan data

Implementasi Sistem penilaian ujian esai ini dimulai dengan tahap persiapan data yang akan diproses. Data ini mencakup soal dan kunci jawaban dari mata pelajaran Sejarah di SMK intensif baitussalam, serta jawaban dari siswa yang akan digunakan sebagai data awal untuk menguji akurasi penilaian esai otomatis dengan menggunakan algoritma *Cosine Similarity*. Berikut adalah data untuk soal esai dalam mata pelajaran sejarah beserta kunci jawaban yang ditunjukkan pada tabel 2.

TABEL II
 DATA SOAL DAN KUNCI JAWABAN

No	Soal	Kunci Jawaban	Bobot
1	Jelaskan peran penting Sumpah Pemuda pada tahun 1928 dalam upaya memperkuat persatuan bangsa Indonesia?	Sumpah Pemuda Memperkuat persatuan dengan mengikrarkan satu tanah air, satu bangsa, dan satu bahasa, yaitu Indonesia.	35
2	Sebutkan dampak dari penjajahan Belanda terhadap sistem ekonomi tradisional masyarakat Indonesia?	Penjajahan Belanda mengakibatkan sistem ekonomi tradisional tergeser oleh sistem tanam paksa, eksploitasi sumber daya alam, dan penghapusan otonomi ekonomi lokal.	35
3	Apa tujuan utama dari Proklamasi Kemerdekaan Indonesia pada 17 Agustus 1945?	Tujuan utamanya adalah untuk menyatakan kemerdekaan bangsa Indonesia dari penjajahan serta menunjukkan kepada dunia bahwa Indonesia telah merdeka dan berdaulat.	30

Terdapat 3 data jawaban siswa yang digunakan dalam proses penilaian otomatis pada penelitian ini. Pada tabel 3,4 dan 5, menunjukkan 3 jawaban esai tersebut yang menjadi sampel untuk evaluasi sistem.

TABEL III
 DATA JAWABAN SISWA SOAL NOMOR 1

Siswa	Jawaban soal nomor 1
S1	Sumpah Pemuda membantu menyatukan rakyat Indonesia dengan memperjuangkan satu tanah air, bangsa, dan bahasa. Ini membuat semangat persatuan semakin kuat untuk melawan penjajah.
S2	Peran Sumpah Pemuda adalah menyatukan pemuda-pemudi dari berbagai daerah untuk bersatu dalam satu tanah air, satu bangsa, dan satu bahasa, yaitu Indonesia.

S3 Sumpah Pemuda berperan memperkuat persatuan bangsa Indonesia dengan menekankan satu tanah air, bangsa, dan bahasa Indonesia sebagai semangat perjuangan melawan penjajahan

TABEL IV
 DATA JAWABAN SISWA SOAL NOMOR 2

Siswa	Jawaban soal nomor 2
S1	Penjajahan Belanda membuat ekonomi tradisional terganggu, seperti dengan adanya tanam paksa dan penguasaan sumber daya oleh Belanda.
S2	Dampaknya adalah ekonomi tradisional masyarakat tergantikan oleh sistem tanam paksa dan eksploitasi sumber daya yang dilakukan Belanda.
S3	Penjajahan Belanda memaksa masyarakat meninggalkan sistem ekonomi tradisional karena ada tanam paksa dan pengambilan hasil bumi oleh penjajah.

TABEL V
 DATA JAWABAN SISWA SOAL NOMOR 3

Siswa	Jawaban soal nomor 3
S1	Proklamasi Kemerdekaan Bertujuan Untuk Menyatakan Bahwa Indonesia Sudah Merdeka Dan Tidak Dijajah Lagi.
S2	Tujuan Utamanya Adalah Agar Indonesia Merdeka Dan Menjadi Negara Yang Bebas Dari Penjajahan.
S3	Proklamasi Kemerdekaan Bertujuan Untuk Menyatakan Bahwa Bangsa Indonesia Sudah Merdeka Dan Berdaulat Sepenuhnya.

B. Text preprocessing

Tahap awal yang dilakukan adalah *text preprocessing* untuk mengubah teks tidak terstruktur menjadi format yang lebih terorganisir. Berikut adalah tabel-tabel hasil dari proses *text preprocessing* dari semua jawaban siswa.

1. Cleaning

Proses mengkonversi teks ke format standar dengan menghilangkan karakter khusus seperti simbol "@", "%", "\$", "!", "?" dan sejenisnya, akan dihapus guna meningkatkan keakuratan hasil pengolahan data.

TABEL VI
 CLEANING JAWABAN SISWA SOAL NOMOR 1

Siswa	Jawaban nomor 1
S1	Sumpah Pemuda Membantu Menyatukan Rakyat Indonesia Dengan Memperjuangkan Satu Tanah Air Bangsa Dan Bahasa Ini Membuat Semangat Persatuan Semakin Kuat Untuk Melawan Penjajah.
S2	Peran Sumpah Pemuda Adalah Menyatukan Pemuda-Pemudi Dari Berbagai Daerah Untuk Bersatu Dalam Satu Tanah Air Satu Bangsa Dan Satu Bahasa Yaitu Indonesia
S3	Sumpah Pemuda Berperan Memperkuat Persatuan Bangsa Indonesia Dengan Menekankan Satu Tanah Air Bangsa Dan Bahasa Indonesia Sebagai Semangat Perjuangan Melawan Penjajahan.

TABEL VII
 CLEANING JAWABAN SISWA SOAL NOMOR 2

Siswa	Jawaban nomor 2
S1	Penjajahan Belanda Membuat Ekonomi Tradisional Terganggu Seperti Dengan Adanya Tanam Paksa Dan Penguasaan Sumber Daya Oleh Belanda
S2	Dampaknya Adalah Ekonomi Tradisional Masyarakat Tergantikan Oleh Sistem Tanam Paksa Dan Eksploitasi Sumber Daya Yang Dilakukan Belanda
S3	Penjajahan Belanda Memaksa Masyarakat Meninggalkan Sistem Ekonomi Tradisional Karena Ada Tanam Paksa Dan Pengambilan Hasil Bumi Oleh Penjajah

TABEL VIII
 CLEANING JAWABAN SISWA SOAL NOMOR 3

Siswa	Jawaban nomor 3
S1	Proklamasi Kemerdekaan Bertujuan Untuk Menyatakan Bahwa Indonesia Sudah Merdeka Dan Tidak Dijajah Lagi
S2	Tujuan Utamanya Adalah Agar Indonesia Merdeka Dan Menjadi Negara Yang Bebas Dari Penjajahan
S3	Proklamasi Kemerdekaan Bertujuan Untuk Menyatakan Bahwa Bangsa Indonesia Sudah Merdeka Dan Berdaulat Sepenuhnya

Tabel 6, 7 dan 8 merupakan hasil dari proses *cleaning* pada jawaban siswa yang akan dilanjut dengan proses *case folding*.

2. Case Folding

Selanjutnya yaitu proses untuk mengkonversi teks ke format standar dengan mengubah huruf besar menjadi huruf besar kecil yang di tunjukan pada tabel 9, 10 dan 11 sebagai berikut :

TABEL IX
 CASE FOLDING JAWABAN SISWA SOAL NOMOR 1

Siswa	Jawaban nomor 1
S1	sumpah pemuda membantu menyatukan rakyat indonesia dengan memperjuangkan satu tanah air bangsa dan bahasa ini membuat semangat persatuan semakin kuat untuk melawan penjajah

S2	peran sumpah pemuda adalah menyatukan pemuda-pemudi dari berbagai daerah untuk bersatu dalam satu tanah air satu bangsa dan satu Bahasa yaitu indonesia
S3	sumpah pemuda berperan memperkuat persatuan bangsa indonesia dengan menekankan satu tanah air bangsa dan bahasa indonesia sebagai semangat perjuangan melawan penjajahan

TABEL X
 CASE FOLDING JAWABAN SISWA SOAL NOMOR 2

Siswa	Jawaban nomor 2
S1	penjajahan belanda membuat ekonomi tradisional terganggu, seperti dengan adanya tanam paksa dan penguasaan sumber daya oleh belanda
S2	dampaknya adalah ekonomi tradisional masyarakat tergantikan oleh sistem tanam paksa dan eksploitasi sumber daya yang dilakukan belanda
S3	penjajahan belanda memaksa masyarakat meninggalkan sistem ekonomi tradisional karena ada tanam paksa dan pengambilan hasil bumi oleh penjajah

TABEL XI
 CASE FOLDING JAWABAN SISWA SOAL NOMOR 3

Siswa	Jawaban nomor 3
S1	proklamasi kemerdekaan bertujuan untuk menyatakan bahwa indonesia sudah merdeka dan tidak dijajah lagi
S2	tujuan utamanya adalah agar indonesia merdeka dan menjadi negara yang bebas dari penjajahan
S3	proklamasi kemerdekaan bertujuan untuk menyatakan bahwa bangsa indonesia sudah merdeka dan berdaulat sepenuhnya

Dari tabel 9, 10 dan 10, merupakan hasil dari proses *case folding* pada jawaban siswa yang akan dilanjut proses *tokenization*.

3. Tokenization

Proses memisahkan teks menjadi unit-unit yang lebih kecil, seperti kata-kata, frasa, atau token. Proses tersebut merupakan proses dalam pemrosesan teks. Misalnya, kalimat “Saya sedang mempelajari teori pemrograman” akan di pisahkan menjadi token “saya”, ”sedang”, ”mempelajari”, ”teori”, ”pemrograman”.

TABEL XII
 TOKENIZATION JAWABAN SISWA SOAL NOMOR 1

Siswa	Jawaban nomor 1
S1	sumpah, pemuda, membantu, menyatukan, rakyat, indonesia, memperjuangkan, satu, tanah, air, bangsa, bahasa, membuat, semangat, persatuan, semakin, kuat, melawan, penjajah
S2	peran, sumpah, pemuda, menyatukan, pemuda-pemudi, berbagai, daerah, bersatu, satu, tanah, air, satu, bangsa, satu, bahasa, indonesia
S3	sumpah, pemuda, berperan, memperkuat, persatuan, bangsa, indonesia, menekankan, satu, tanah, air, bangsa, bahasa, indonesia, semangat, perjuangan, melawan, penjajahan

TABEL XIII
 TOKENIZATION JAWABAN SISWA SOAL NOMOR 2

Siswa	Jawaban nomor 2
S1	penjajahan, belanda, membuat, ekonomi, tradisional, terganggu, adanya, tanam, paksa, penguasaan, sumber, daya, belanda
S2	dampaknya, ekonomi, tradisional, masyarakat, tergantikan, sistem, tanam, paksa, eksploitasi, sumber, daya, dilakukan, belanda
S3	penjajahan, belanda, memaksa, masyarakat, meninggalkan, sistem, ekonomi, tradisional, tanam, paksa, pengambilan, hasil, bumi, penjajah

TABEL XIV
 TOKENIZATION JAWABAN SISWA SOAL NOMOR 3

Siswa	Jawaban nomor 3
S1	proklamasi, kemerdekaan, bertujuan, menyatakan, indonesia, merdeka, dijajah
S2	tujuan, utamanya, indonesia, merdeka, menjadi, negara, bebas, penjajahan
S3	proklamasi, kemerdekaan, bertujuan, menyatakan, bangsa, indonesia, merdeka, berdaulat, sepenuhnya

Tabel 12, 13 dan 14, merupakan hasil dari proses *tokenization* pada jawaban siswa yang akan dilanjut proses *stopword removal*.

4. Stopword Removal

Teknik untuk menghilangkan kata-kata yang dianggap tidak penting, seperti kata sambung, kata depan, kata ganti, atau kata-kata yang tidak relevan untuk analisis sentimen, sehingga perhitungan dapat berkonsentrasi pada kata-kata yang lebih relevan.

TABEL XV
 STOPWORD REMOVAL JAWABAN SISWA SOAL NOMOR 1

SISWA	Jawaban nomor 1
S1	sumpah, pemuda, membantu, menyatukan, rakyat, indonesia, memperjuangkan, satu, tanah, air, bangsa, bahasa, membuat, semangat, persatuan, semakin, kuat, melawan, penjajah
S2	peran, sumpah, pemuda, menyatukan, pemuda-pemudi, berbagai, daerah, bersatu, satu, tanah, air, satu, bangsa, satu, bahasa, indonesia

S3	sumpah, pemuda, berperan, memperkuat, persatuan, bangsa, indonesia, menekankan, satu, tanah, air, bangsa, bahasa, indonesia, semangat, perjuangan, melawan, penjajahan
----	--

TABEL XVI
 STOPWORD REMOVAL JAWABAN SISWA SOAL NOMOR 2

SISWA	Jawaban nomor 2
S1	penjajahan, belanda, membuat, ekonomi, tradisional, terganggu, adanya, tanam, paksa, penguasaan, sumber, daya, belanda
S2	dampaknya, ekonomi, tradisional, masyarakat, tergantikan, sistem, tanam, paksa, eksploitasi, sumber, daya, dilakukan, belanda
S3	penjajahan, belanda, memaksa, masyarakat, meninggalkan, sistem, ekonomi, tradisional, tanam, paksa, pengambilan, hasil, bumi, penjajah

TABEL XVII
 STOPWORD REMOVAL JAWABAN SISWA SOAL NOMOR 3

SISWA	Jawaban nomor 3
S1	proklamasi, kemerdekaan, bertujuan, menyatakan, indonesia, merdeka, dijajah
S2	tujuan, utamanya, indonesia, merdeka, menjadi, negara, bebas, penjajahan
S3	proklamasi, kemerdekaan, bertujuan, menyatakan, bangsa, indonesia, merdeka, berdaulat, sepenuhnya

Tabel 15, 16 dan 17, merupakan hasil dari proses *stopword removal* pada jawaban siswa yang akan dilanjut dengan proses *stemming*.

5. Stemming

Menghapus imbuhan kata untuk menghasilkan bentuk kata dasar, baik berupa awalan ataupun akhiran. Pada dasarnya proses yang dilakukan ini menggunakan basis pengetahuan yang disebut *basicword list*.

TABEL XVIII
 STEMMING JAWABAN SISWA SOAL NOMOR 1

SISWA	Jawaban nomor 1
S1	sumpah, pemuda, bantu, satu, rakyat, indonesia, juang, satu, tanah, air, bangsa, bahasa, buat, semangat, satu, makin, kuat, lawan, jajah
S2	peran, sumpah, pemuda, menyatukan, pemuda pemudi, berbagai, daerah, bersatu, satu, tanah, air, satu, bangsa, satu, bahasa, indonesia
S3	sumpah, pemuda, peran, kuat, satu, bangsa, indonesia, tekan, satu, tanah, air, bangsa, bahasa, indonesia, semangat, juang, lawan, jajah

TABEL XIX
 STEMMING JAWABAN SISWA SOAL NOMOR 2

SISWA	Jawaban nomor 2
S1	jajah, belanda, buat, ekonomi, tradisional, ganggu, ada, tanam, paksa, kuasa, sumber, daya, belanda
S2	dampak, ekonomi, tradisional, masyarakat, ganti, sistem, tanam, paksa, eksploitasi, sumber, daya, laku, belanda
S3	jajah, belanda, paksa, masyarakat, tinggal, sistem, ekonomi, tradisional, tanam, paksa, ambil, hasil, bumi, jajah

TABEL XX
 STEMMING JAWABAN SISWA SOAL NOMOR 3

SISWA	Jawaban nomor 3
S1	proklamasi, merdeka, tuju, nyata, indonesia, merdeka, jajah
S2	tuju, utama, indonesia, merdeka, jadi, negara, bebas, jajah
S3	proklamasi, merdeka, tuju, nyata, bangsa, indonesia, merdeka, daulat, sepenuh

Tabel 18, 19 dan 20, merupakan hasil dari proses *stemming* yang merupakan proses terakhir dari teks *preprocessing* pada jawaban siswa, selanjutnya jawaban akan di proses dengan pembobotan kata untuk menghitung seberapa sering kata pada jawaban muncul dan dihitung menggunakan algoritma *cosine similarity*.

C. Implementasi Algoritma

Setelah melewati tahap *preprocessing*, proses selanjutnya adalah melakukan penilaian otomatis menggunakan algoritma. Langkah pertama dalam tahap ini adalah menghitung bobot kata menggunakan *Term Frequency*, yaitu menghitung seberapa sering suatu kata muncul dalam sebuah jawaban atau dokumen. Nilai frekuensi ini kemudian digunakan dalam perhitungan menggunakan algoritma *Cosine Similarity* untuk mengukur tingkat kemiripan antara jawaban siswa dengan kunci jawaban. Semakin mendekati angka 1 hasil *cosine similarity* yang diperoleh, maka semakin tinggi tingkat kesesuaian jawaban siswa terhadap kunci jawaban. Selanjutnya, skor hasil perhitungan *cosine similarity* dikalikan dengan bobot nilai masing-masing soal. Seluruh hasil perkalian tersebut dijumlahkan untuk menghasilkan nilai akhir dari jawaban yang diberikan oleh siswa.

1) *Term Frequency Inverse Document Frequency*

Perhitungan *term frequency* dilakukan dengan menggabungkan semua term jawaban yang unik, kemudian menghitung frekuensi setiap kata dan menggunakan nilai bobot *term frequency* untuk setiap kata. selanjutnya yaitu *Inverse Document Frequency (IDF)* bertujuan untuk mengukur tingkat keunikan suatu term dalam kumpulan dokumen. Jika sebuah kata muncul di banyak dokumen, maka nilai IDF-nya akan rendah, menandakan bahwa kata tersebut kurang informatif. Sebaliknya, jika sebuah kata hanya muncul di sedikit dokumen, maka nilai IDF-nya tinggi.

TABEL XXI
 TERM FREQUENCY INVERSE DOCUMENT FREQUENCY JAWABAN SISWA SOAL NOMOR 1

term(t)	tf					IDF	KJ	W= tf*(idf+1)		
	KJ	JS1	JS2	JS3	DF			JS1	JS2	JS3
sumpah	1	1	1	1	4	0	1	1	1	1
pemuda	1	1	2	1	5	-0,0969	0,9031	0,9031	1,8062	0,9031
kuat	1	1	0	1	3	0,1249	1,1249	1,1249	0	1,1249
satu	3	3	4	2	12	-0,4771	1,5686	1,5686	2,0915	1,0458
ikrar	1	0	0	0	1	0,6021	1,6021	0	0	0
tanah	1	1	0	1	3	0,1249	1,1249	1,1249	0	1,1249
air	1	1	1	1	4	0	1,0000	1,0000	1,0000	1,0000
bangsa	1	1	1	2	5	-0,0969	0,9031	0,9031	0,9031	1,8062
bahasa	1	1	1	1	4	0	1,0000	1,0000	1,0000	1,0000
indonesia	1	1	1	2	5	-0,0969	0,9031	0,9031	0,9031	1,8062
bantu	0	1	0	0	1	0,6021	0	1,6021	0	0
rakyat	0	1	0	0	1	0,6021	0	1,6021	0	0
juang	0	1	0	1	2	0,3010	0	1,3010	0	1,3010
buat	0	1	0	0	1	0,6021	0	1,6021	0	0
semangat	0	1	0	1	2	0,3010	0	1,3010	0	1,3010
makin	0	1	0	0	1	0,6021	0	1,6021	0	0
lawan	0	1	0	1	2	0,3010	0	1,3010	0	1,3010
jajah	0	1	0	1	2	0,3010	0	1,3010	0	1,3010
peran	0	0	1	1	2	0,3010	0	0	1,3010	1,3010
pemudi	0	0	1	0	1	0,6021	0	0	1,6021	0
bagai	0	0	1	0	1	0,6021	0	0	1,6021	0
daerah	0	0	1	0	1	0,6021	0	0	1,6021	0
tekan	0	0	0	1	1	0,6021	0	0	0	1,6021

TABEL XXII
 TREM FREQUENCY INVERSE DOCUMENT FREQUENCY JAWABAN SISWA SOAL NOMOR 2

term(t)	tf					IDF	KJ	W= tf*(idf+1)		
	KJ	JS1	JS2	JS3	DF			JS1	JS2	JS3
jajah	1	0	0	1	2	0,3010	1,3010	0	0	1,3010
belanda	1	1	1	1	4	0	1,0000	1,0000	1,0000	1,0000
akibat	1	0	0	0	1	0,6021	1,6021	0	0	0
sistem	2	0	1	1	4	0	2,0000	0,0000	1,0000	1,0000
ekonomi	2	1	1	1	5	-0,0969	1,8062	0,9031	0,9031	0,9031
tradisional	1	1	1	1	4	0	1,0000	1,0000	1,0000	1,0000
geser	1	0	0	0	1	0,6021	1,6021	0	0	0
tanam	1	1	1	1	4	0	1,0000	1,0000	1,0000	1,0000
paksa	1	1	1	1	4	0	1,0000	1,0000	1,0000	1,0000
eksploitasi	1	0	1	0	2	0,3010	1,3010	0	1,3010	0
sumber	1	1	1	0	3	0,1249	1,1249	1,1249	1,1249	0
daya	1	1	1	0	3	0,1249	1,1249	1,1249	1,1249	0
alam	1	0	0	0	1	0,6021	1,6021	0	0	0
hapus	1	0	0	0	1	0,6021	1,6021	0	0	0
otonomi	1	0	0	0	1	0,6021	1,6021	0	0	0
lokal	1	0	0	0	1	0,6021	1,6021	0	0	0
ganggu	0	1	0	0	1	0,6021	0	1,6021	0	0
kuasa	0	1	0	0	1	0,6021	0	1,6021	0	0
dampak	0	0	1	0	1	0,6021	0	0	1,6021	0
masyarakat	0	0	1	1	2	0,3010	0	0	1,3010	1,3010
ganti	0	0	1	0	1	0,6021	0	0	1,6021	0
laku	0	0	1	0	1	0,6021	0	0	1,6021	0
tinggal	0	0	0	1	1	0,6021	0	0	0	1,6021
ambil	0	0	0	1	1	0,6021	0	0	0	1,6021
bumi	0	0	0	1	1	0,6021	0	0	0	1,6021

TABEL XXIII
 TREM FREQUENCY INVERSE DOCUMENT FREQUENCY JAWABAN SISWA SOAL NOMOR 3

term(t)	tf					IDF	KJ	W= tf*(idf+1)		
	KJ	JS1	JS2	JS3	DF			JS1	JS2	JS3
tuju	1	1	1	1	4	0	1,0000	1,0000	1,0000	1,0000
utama	1	0	1	0	2	0,3010	1,3010	0	1,3010	0
nyata	1	1	0	1	3	0,1249	1,1249	1,1249	0	1,1249
merdeka	2	1	1	1	5	0,0969	1,8062	0,9031	0,9031	0,9031

bangsa	1	1	0	2	4	0	1,0000	1,0000	0	2,0000
indonesia	2	1	1	1	5	0,0969	1,8062	0,9031	0,9031	0,9031
jajah	1	1	1	0	3	0,1249	1,1249	1,1249	1,1249	0
tunjuk	1	0	0	0	1	0,6021	1,6021	0	0	0
dunia	1	0	0	0	1	0,6021	1,6021	0	0	0
daulat	1	0	0	1	2	0,3010	1,3010	0	0	1,3010
proklmasi	0	1	0	1	2	0,3010	0	1,3010	0	1,3010
jadi	0	0	1	0	1	0,6021	0	0	1,6021	0
negara	0	0	1	0	1	0,6021	0	0	1,6021	0
bebas	0	0	1	0	1	0,6021	0	0	1,6021	0
sepenuh	0	0	0	1	1	0,6021	0	0	0	1,6021

Tabel 21, 22 dan 23, merupakan hasil dari perhitungan dari *TF-IDF* dari kunci jawaban dan jawaban siswa untuk selanjutnya akan dihitung menggunakan *cosine similarity*.

2) Cosine Similarity

Setelah dilakukan pembobotan *Term Frequency Inverse Document Frequency* pada kunci jawaban dan data jawaban setiap soal, tahap berikutnya adalah menghitung tingkat kesamaan antara dokumen jawaban dengan kunci jawaban. Proses ini menggunakan algoritma *Cosine Similarity*, yang berfungsi untuk mengukur tingkat kesamaan antara dua dokumen yang dinyatakan dalam bentuk dua vektor. Berikut ini merupakan hasil perhitungan nilai *Cosine Similarity* untuk setiap pasangan kunci jawaban dan data jawaban pada setiap soal.

TABEL XXIV
 PERHITUNGAN COSINE SIMILARITY JAWABAN SISWA SOAL NOMOR 1

Siswa (S)	$\sqrt{\Sigma(J)^2}$	$\sqrt{\Sigma(S)^2}$	$\sqrt{\Sigma(J)^2} \cdot \sqrt{\Sigma(S)^2}$	Cosine Similarity
S1	3,4641	3,687817	12,7750	0,93934
S2	3,4641	3,87298	13,4164	0,89443
S3	3,4641	3,92428	13,5941	0,88274

TABEL XXV
 PERHITUNGAN COSINE SIMILARITY JAWABAN SISWA SOAL NOMOR 2

Siswa (S)	$\sqrt{\Sigma(J)^2}$	$\sqrt{\Sigma(S)^2}$	$\sqrt{\Sigma(J)^2} \cdot \sqrt{\Sigma(S)^2}$	Cosine Similarity
S1	4,24264	1,974841	8,3785	0,4655
S2	4,24264	2,82842	12,0000	0,6667
S3	4,24264	3,162277	13,4164	0,5094

TABEL XXVI
 PERHITUNGAN COSINE SIMILARITY JAWABAN SISWA SOAL NOMOR 3

Siswa (S)	$\sqrt{\Sigma(J)^2}$	$\sqrt{\Sigma(S)^2}$	$\sqrt{\Sigma(J)^2} \cdot \sqrt{\Sigma(S)^2}$	Cosine Similarity
S1	3,742	5,196	19,442	0,617215
S2	3,742	5,523	20,664	0,580722
S3	3,742	4,593	17,187	0,698206

Tabel 24, 25 dan 26, menunjukkan proses perhitungan nilai *Cosine Similarity* untuk soal-soal nomor 1 hingga nomor 3 berdasarkan data jawaban dari masing-masing siswa. Setelah memperoleh nilai *Cosine Similarity*, langkah terakhir yang dilakukan untuk mengetahui hasil penilaian adalah mengalikan nilai *Cosine Similarity* dan bobot nilai dari masing-masing soal yang telah ditetapkan sebelumnya.

3) Hasil Penilaian

Setelah mendapatkan hasil dari perhitungan *Cosine Similarity* dari setiap soal maka nilai *Cosine Similarity* tersebut dikalikan dengan bobot nilai setiap soal, bobot nilai untuk masing-masing soal dapat dilihat kembali pada tabel 2 soal dan kunci jawaban diatas. Hasil dari perkalian tersebut dijumlahkan semua untuk memperoleh nilai akhir dari setiap siswa. Hasil penilaian akhir yang didapat dari hasil pengolahan data *Natural Language Processing* dengan pendekatan *Term frequency* dan Algoritma *Cosine Similarity* ditunjukkan pada tabel berikut 27 dan 28.

TABEL XXVII
 HASIL TINGKAT KEMIRIPAN COSINE SIMILARITY

Siswa	Cosim Soal 1	Cosim Soal 2	Cosim Soal 3	Similarity (J1)%	Similarity (J2)%	Similarity (J3)%
S1	0,93934	0,4655	0,6172	94%	47%	62%

S2	0,89443	0,6667	0,5807	89%	67%	58%
S3	0,88274	0,5094	0,6982	88%	51%	70%

TABEL XXVIII
 HASIL TINGKAT PENILAIAN COSINE SIMILARITY

Siswa	Cosim soal 1	Cosim Soal 2	Cosim Soal 3	N1	N2	N3	Hasil = N1+N2+N3
S1	0,93934	0,4655	0,617215	32,87680	16,292	18,5164	67,68
S2	0,89443	0,6667	0,580722	31,30499	23,333	17,4217	72,05
S3	0,88274	0,5094	0,698206	30,89576	17,829	20,9462	69,67

D. Hasil Evaluasi Sistem Dan Penilaian Guru Manual

Untuk mengevaluasi efektivitas sistem penilaian esai otomatis, dilakukan perbandingan hasil antara skor yang diberikan oleh sistem dan skor yang diberikan oleh guru secara manual. Perbandingan ini bertujuan untuk menilai tingkat akurasi dan konsistensi sistem dalam menilai jawaban siswa, sebagai label referensi peneliti mengambil 10 data hasil penilaian dari sistem dan penilaian manual dari guru. Dalam penelitian ini, *threshold* adalah teknik yang digunakan untuk menetapkan nilai batas dalam menentukan terjadinya suatu kondisi atau kejadian tertentu, dilakukan klasifikasi biner dengan menggunakan nilai ambang batas *threshold* sebesar 75 [23]. Aturan klasifikasi adalah Jika skor lebih dari ≥ 75 , maka diklasifikasikan sebagai 1 (bagus). Jika skor kurang dari < 75 , maka diklasifikasikan sebagai 0 (kurang) perbandingan nilai program dan nilai manual oleh guru dapat dilihat pada tabel 29.

TABEL XXIX
 KONVERSI SKOR MENJADI KELAS KLASIFIKASI

Siswa	Skor Program	Skor Guru	Prediksi (Program)	Aktual (Guru)
1	78	80	1	1
2	83	85	1	1
3	74	79	0	1
4	79	80	1	1
5	78	80	1	1
6	81	85	1	1
7	77	80	1	1
8	73	77	0	1
9	80	82	1	1
10	79	81	1	1

Berdasarkan hasil klasifikasi pada tabel 29, dilakukan perhitungan jumlah *True Positive* (TP), *False Positive* (FP), *False Negative* (FN), dan *True Negative* (TN) seperti tabel 30 dan 31 sebagai berikut:

TABEL XXX
 HASIL CONFUSION MATRIX

Actual value	Predicted value	
	Positive (1)	Negative(0)
Positive(1)	TP = 8	FN = 2
Negative(0)	FP = 0	TN = 0

TABEL XXXI
 HASIL PERHITUNGAN MATRIX EVALUASI

Metrik	Rumus	Hasil
Akurasi	$(TP + TN) / (TP + TN + FP + FN)$	$8 / 10 = 0.80$
Presisi	$TP / (TP + FP)$	$8 / 8 = 1.00$
Recall	$TP / (TP + FN)$	$8 / (8 + 2) = 0.80$
F1-score	$2 \times (Presisi \times Recall) / (Presisi + Recall)$	$2 \times (1 \times 0.80) / (1 + 0.80) = 0.89$

Hasil evaluasi yang ditunjukkan pada tabel 31 bahwa sistem memiliki akurasi sebesar 80%, dengan presisi sempurna 100% karena seluruh prediksi positif yang dilakukan oleh sistem tepat sasaran tidak terdapat *False Positive*. Namun, terdapat dua kasus di mana sistem gagal mengklasifikasikan skor yang seharusnya masuk kategori bagus *False Negative*, sehingga nilai *recall* mencapai 80%. Metrik *F1-score* yang merupakan rata-rata harmonis dari *presisi* dan *recall* menghasilkan nilai 89%, yang mencerminkan kinerja sistem yang baik

dan seimbang. Dari hasil perhitungan sistem penilaian otomatis ini rata-rata selisih antara penilaian manual dan otomatis adalah kurang lebih 3 *point*, yang menunjukkan bahwa sistem memiliki tingkat akurasi yang tinggi dan cukup konsisten dalam menilai esai siswa. Hal ini menunjukkan bahwa algoritma yang digunakan berhasil mengidentifikasi kemiripan makna antara jawaban siswa dan kunci secara cukup presisi.

Berdasarkan hasil evaluasi dari penilaian otomatis dari sistem dan penilaian guru secara manual, diketahui bahwa sistem memberikan hasil yang cukup dekat dengan penilaian manual oleh guru. Adapun keunggulan sistem ini adalah mampu mempercepat proses penilaian esai secara otomatis, mengurangi subjektivitas guru, dan memberikan hasil yang konsisten. Dengan pendekatan *cosine similarity* cukup efektif menilai kemiripan teks antara jawaban siswa dan kunci jawaban, terutama jika siswa menggunakan kata-kata yang relevan meskipun dengan susunan berbeda. Adapun kelemahan dari sistem ini belum bisa memahami makna konteks secara mendalam. Jika siswa menggunakan bahasa yang sangat berbeda dari kunci jawaban, hasil penilaiannya bisa kurang akurat.

Penggunaan sistem penilaian otomatis ini berkontribusi dalam meningkatkan efektivitas proses pembelajaran. Guru tidak lagi menghabiskan banyak waktu untuk mengoreksi jawaban esai secara manual, sehingga waktu tersebut dapat dialihkan untuk memberikan umpan balik atau melakukan pembelajaran remedial bagi siswa yang nilainya kurang. Bagi siswa, sistem ini memberikan hasil penilaian yang lebih cepat, sehingga mereka bisa segera mengetahui kelemahan dalam jawaban mereka dan memperbaikinya. Hal ini membantu siswa untuk lebih aktif dalam proses refleksi pembelajaran. Selain itu, guru dapat menggunakan hasil evaluasi sebagai data awal untuk menganalisis pola pemahaman siswa terhadap materi, serta membuat strategi pengajaran lanjutan yang lebih tepat sasaran.

IV. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa sistem penilaian otomatis menggunakan teknik *text preprocessing*, *term frequency*, dan *cosine similarity* mampu memberikan hasil klasifikasi yang cukup akurat jika dibandingkan dengan penilaian guru secara manual. Evaluasi dilakukan dengan membandingkan skor hasil program dengan skor guru menggunakan pendekatan klasifikasi biner dengan *threshold 75*. Hasil evaluasi menunjukkan bahwa sistem menghasilkan akurasi sebesar 80%, presisi 100%, recall 80%, dan F1-score sebesar 89%. Hal ini menunjukkan bahwa sistem cukup efektif dalam mengenali esai yang berkualitas tinggi, meskipun masih terdapat beberapa kasus esai yang tidak terklasifikasi secara tepat. Secara keseluruhan, sistem yang dikembangkan dapat menjadi alat bantu yang potensial dalam meringankan beban guru dalam proses penilaian esai, terutama pada jumlah peserta didik yang banyak. Namun demikian, pengembangan lebih lanjut diperlukan, seperti penyesuaian bobot kata, peningkatan *preprocessing*, dan uji coba pada berbagai jenis soal esai untuk meningkatkan akurasi sistem secara keseluruhan.

DAFTAR PUSTAKA

- [1] Imam Mufiid, S. Lestanti, and N. Kholila, "Aplikasi Penilaian Jawaban Esai Otomatis Menggunakan Metode Synonym Recognition Dan Cosine Similarity Berbasis Web," *J. Mnemon.*, vol. 4, no. 2, pp. 31–37, 2021, doi: 10.36040/mnemonic.v4i2.4067.
- [2] *Kelebihan dan Keterbatasan dari Bentuk-Bentuk Penilaian Tes Tertulis.*
- [3] F. E. Kurniawati and W. Mega Pradnya, "Implementasi Algoritma Winnowing Pada Sistem Penilaian Otomatis Jawaban Esai Pada Ujian Online Berbasis Web," *J. Tek. Komput. AMIK BSI*, vol. VI, no. 2, 2020, doi: 10.31294/jtk.v4i2.
- [4] Y. Witdianti, *Mengenal Higher Order Thinking Skills Lebih Dalam*. J. Rajawali, G. Elang 6, No 3, Deron, Sardono Harjo, Ngaglik, Sleman. Jl. Kaliurang Km. 9.3 - Yogyakarta 55581: Deepublish, 2023.
- [5] E. L. Amalia¹, A. J. Jumadi, I. A. Mashudi², W. Wibowo⁴, P. N. Malang, and P. Korespondensi, "ANALISIS METODE COSINE SIMILARITY PADA APLIKASI UJIAN ONLINE ESAI OTOMATIS (STUDI KASUS JTI POLINEMA) COSINE SIMILARITY METHOD ANALYSIS ON AUTOMATIC ESAI ONLINE TEST APPLICATION", doi: 10.25126/jtiik.202184356.
- [6] J. R. Triosaputra, A. Sanjaya, J. Sahertian, U. Nusantara, and P. Kediri, "Penilaian Otomatis Cerdas Cermat Menggunakan Basis Data Sinonim Kata Dan Cosine Similarity," Online, 2024.
- [7] N. L. Kinanti and A. Qoiriah, "Sistem Penilaian Otomatis Jawaban Esai Bahasa Indonesia Berdasarkan Kemiripan Kalimat Menggunakan Syntactic- Semantic Similarity," *JINACS (Journal Informatics Comput. Sci.)*, vol. 2, no. 2, pp. 136–143, 2020.
- [8] S. Rifki, *ARTIFICIAL INTELLIGENCE: Teori dan Penerapan AI di berbagai bidang*. Pertama. Jambi: PT. Sonpedia publishing Indonesia, 2024.
- [9] N. Nur Azizah, I. Purnamasari, and S. Prangga, "Pengelompokan Judul Laporan Skripsi Berbasis Text Mining dengan Metode Fuzzy K-Means," *METIK J.*, vol. 8, no. 1, pp. 18–23, Jun. 2024, doi: 10.47002/metik.v8i1.808.
- [10] A. Apriani, H. Zakiyudin, and K. Marzuki, "Penerapan Algoritma Cosine Similarity dan Pembobotan TF-IDF System Penerimaan Mahasiswa Baru pada Kampus Swasta," *J. Bumigora Inf. Technol.*, vol. 3, no. 1, pp. 19–27, Jul. 2021, doi: 10.30812/bite.v3i1.1110.
- [11] M. Febima, L. Magdalena, M. Asfi, M. Hatta, and R. Fahrudin, "Implementasi Optimasi NLP dan KNN untuk User Review Aplikasi SAMPEAN Cirebon," pp. 162–168.
- [12] A. A. Syam, G. H. M, A. Salim, D. F. Suriyanto, and M. F. B, "Analisis teknik preprocessing pada sentimen masyarakat terkait konflik israel-palestina menggunakan support vector machine," vol. 9, no. 3, pp. 1464–1472, 2024.
- [13] U. Khairani, V. Mutiawani, and H. Ahmadian, "Pengaruh Tahapan Preprocessing Terhadap Model Indobert Dan Indobertweet Untuk Mendeteksi

- Emosi Pada Komentar Akun Berita Instagram,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 11, no. 4, pp. 887–894, 2024, doi: 10.25126/jtiik.1148315.
- [14] A. H. Azmi, N. I. Mahardika, R. S. Anugrah, and A. P. Sari, “Journal of Multidisciplinary Inquiry in Science Technology and Educational Research Sistem Rekomendasi Film Berbasis Konten Menggunakan Teknik Cosine Similarity dan TF-IDF”.
- [15] A. Sanjaya, A. B. Setiawan, U. Mahdiyah, I. N. Farida, and A. R. Prasetyo, “Pengukuran Kemiripan Makna Menggunakan Cosine Similarity dan Basis Data Sinonim Kata,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 10, no. 4, pp. 747–752, 2023, doi: 10.25126/jtiik.20241046864.
- [16] A. Sanjaya; A. B. Setiawan; D. P. Pamungkas; I. N. Farida and M. A. D. Widyadara;, “Measuring Meaning Similarity Using TF/IDF and Term Synonym ID,” *2023 6th Int. Conf. Inf. Commun. Technol.*, pp. 206–211, 2023, doi: doi: 10.1109/ICOIACT59844.2023.10455894.
- [17] M. R. Arifuddin, I. A. Rafiq, R. Mubarak, and P. H. Susilo, “Sistem Cerdas Penilaian Ujian Essay Menggunakan Metode Cosine Similarity.”
- [18] E. Putra Pane, R. Hardianto, and W. Choriah, “SISTEM PENILAIAN UJIAN ESAI SECARA OTOMATIS DENGAN ALGORITMA TEXT MINING COSINE SIMILARITY PENUNJANG PEMBELAJARAN,” 2024.
- [19] A. R. Lahitani, “Automated Essay Scoring menggunakan Cosine Similarity pada Penilaian Esai Multi Soal,” *J. Kaji. Ilm.*, vol. 22, no. 2, pp. 107–118, 2022, doi: 10.31599/jki.v22i2.1121.
- [20] R. Rismayani, H. SY, T. Darwansyah, and I. Mansyur, “Implementasi Algoritma Text Mining dan Cosine Similarity untuk Desain Sistem Aspirasi Publik Berbasis Mobile,” *Komputika J. Sist. Komput.*, vol. 11, no. 2, pp. 169–176, 2022, doi: 10.34010/komputika.v11i2.6501.
- [21] F. Rifaldy, Y. Sibaroni, and S. S. Prasetiyowati, “Effectiveness of word2vec and tf-idf in sentiment classification on online investment platforms using support vector machine 1.,” vol. 10, no. 2, pp. 863–874, 2025.
- [22] D. Normawati and S. A. Prayogi, “Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter,” *J. Sains Komput. Inform. (J-SAKTI)*, vol. 5, no. 2, pp. 697–711, 2021.
- [23] Ni luh Putu Trisna Ristanti and Rizky Pradana, “Penggunaan Metode Threshold dalam Pembuatan Sistem Pendeteksi Asap dan Api dengan Berbasis Firebase Dan Android Menggunakan Nodemcu Pada BJ House 77,” *J. Ticom Technol. Inf. Commun.*, vol. 11, no. 1, pp. 44–49, 2022, doi: 10.70309/ticom.v11i1.70.