

Penerapan Algoritma Cosine Similarity dan Term Frequency untuk Penilaian Jawaban Ujian Esai

Martin Hidayat Rihwan^{1*}, Hidola Syamsito²

^{1,2}Program Studi Teknik Informatika, Fakultas Informatika dan Komputer, Universitas Binaniaga Indonesia
email: martin.hidayat.rihwan@gmail.com

*Corresponding Author

ABSTRACT

Exams are one form of educational evaluation to measure students' ability to understand lessons. Essays are one type of exam that requires students to describe the answers using their own words. Essay answer assessment is done by matching the similarity of students' answers with the answer key. The problems that occur during the assessment are the process which takes a relatively long time, there are errors in giving points, and the giving of marks is not quite right. There needs to be a computational engineering approach to make it easier for teachers to give marks using the Term Frequency and Cosine Similarity algorithms. With this algorithm, the essay assessment process is expected to be able to increase the level of accuracy and effectiveness. This assessment process is carried out by measuring the level of similarity between the essay answer text and the answer key text based on the frequency of occurrence of the essay answer word weight and the answer key. the calculations obtained from the algorithm calculation are as follows Similarity to the student's answer key as many as 10 students get scores of 76, 71, 84, 71, 84, 76, 84, 88, 88 and 71. In the developed application, a feasibility test has been carried out by the system expert and the feasibility percentage results are 100%. user test using the PSSUQ questionnaire which is divided into four categories, including Overall of 86.63%, System Usability of 87.67%, Information Quality of 87.25%, and Interface Quality of 80.79% which means this application is "Very Worthy" to use. test results on this algorithm using the F1-Score method which produces an accuracy value of 0.824 which is included in the "good" category.

Keywords: Essay Assessment, Term Frequency Algorithm, Cosine Similarity Algorithm, F1-Score, Application Prototype.

ABSTRAK

Ujian adalah salah satu bentuk evaluasi pendidikan untuk mengukur kemampuan siswa/i dalam memahami pelajaran. Esai merupakan salah satu jenis ujian yang menuntut siswa/i untuk menguraikan jawaban menggunakan kata – katanya sendiri. Penilaian jawaban esai dilakukan dengan cara mencocokkan kemiripan jawaban siswa/i dengan kunci jawaban. Permasalahan yang terjadi selama melakukan penilaian adalah prosesnya yang memakan waktu yang relatif tidak singkat, adanya kesalahan dalam memberikan poin nilai, dan pemberian nilai yang dilakukan kurang tepat. Perlu adanya pendekatan teknik komputasi untuk memudahkan guru dalam memberikan nilai menggunakan algoritma Term Frequency dan Cosine Similarity. Dengan algoritma tersebut, proses penilaian esai diharapkan mampu meningkatkan tingkat ketepatan dan efektivitas. Proses penilaian ini dilakukan dengan mengukur tingkat kemiripan antara teks jawaban esai dengan teks kunci jawaban berdasarkan frekuensi kemunculan bobot kata jawaban esai dan kunci jawaban. perhitungan yang diperoleh dari perhitungan algoritma adalah sebagai berikut Kemiripan dengan kunci jawaban siswa sebanyak 10 siswa mendapatkan nilai sebesar 76, 71, 84, 71, 84, 76, 84, 88, 88 dan 71. Pada aplikasi yang dikembangkan telah dilakukan uji kelayakan dari pihak ahli sistem dan diperoleh hasil persentase kelayakan sebesar 100%. uji pengguna menggunakan kuesioner PSSUQ yang dibagi menjadi empat kategori, di antaranya Secara Keseluruhan sebesar 86.63%, Kegunaan Sistem sebesar 87.67%, Kualitas Informasi sebesar 87.25%, dan Kualitas Antarmuka sebesar 80.79% yang artinya aplikasi ini “Sangat Layak” digunakan. uji hasil pada algoritma ini menggunakan metode F1-Score yang menghasilkan nilai ketepatan sebesar 0.824 yang dimana termasuk kedalam kategori “baik”.

Kata Kunci : Penilaian Esai, Algoritma Term Frequency, Algoritma Cosine Similarity, F1-Score, Prototype Aplikasi.

A. PENDAHULUAN

1. Latar Belakang

Ujian esai telah menjadi salah satu metode evaluasi yang umum digunakan di berbagai tingkat pendidikan dan disiplin ilmu. Metode ini memungkinkan siswa untuk mengekspresikan pemikiran mereka secara mendalam tentang topik tertentu dan menunjukkan pemahaman yang mereka miliki. Pertama-tama, ujian esai memungkinkan guru atau pengajar untuk mengukur pemahaman siswa terhadap materi pelajaran yang telah diajarkan. Dengan memberikan pertanyaan terbuka, ujian esai memungkinkan siswa untuk menunjukkan pemahaman mereka secara menyeluruh daripada hanya mengingat fakta-fakta tertentu. Hal ini memungkinkan siswa untuk menerapkan konsep-konsep yang telah mereka pelajari dalam situasi nyata atau dalam konteks yang berbeda, sehingga memperdalam pemahaman mereka.

Untuk memahami latar belakang penelitian ini lebih lanjut, penting untuk menjelaskan algoritma cosine similarity dan term frequency secara lebih rinci. term frequency (TF) mengukur seberapa sering sebuah kata muncul dalam suatu dokumen dengan membagi jumlah kemunculan kata tersebut dengan total jumlah kata dalam dokumen tersebut. Sementara itu, Cosine similarity adalah metode yang mengukur sejauh mana jawaban siswa mirip dengan kunci jawaban yang benar, mengingat aspek semantik dan konten. Pemilihan algoritma cosine similarity dan term frequency untuk penilaian essay memiliki alasan kuat. Penilaian yang bersifat subjektif dan cenderung memberikan penilaian yang tidak konsisten dapat diatasi dengan penggunaan

algoritma ini. cosine similarity dan term frequency memiliki kerangka kerja yang objektif, yang meminimalkan potensi bias penilaian dan memberikan keadilan dalam.

Penggunaan algoritma Cosine similarity dan Term frequency (TF) dalam penilaian jawaban ujian siswa memiliki beberapa alasan yang mendasar. Pertama, algoritma Cosine similarity digunakan untuk mengukur sejauh mana kesamaan antara jawaban siswa dan kunci jawaban yang benar. Dengan mewakili jawaban sebagai vektor dalam ruang kata-kunci, Cosine similarity memungkinkan evaluasi yang lebih kontekstual daripada metode perbandingan yang sederhana, memperhitungkan seberapa mirip pola kata-kata yang digunakan oleh siswa dengan kunci jawaban. Kedua, penggunaan term frequency dalam penilaian jawaban ujian memberikan pemahaman tentang frekuensi kemunculan kata-kata kunci dalam jawaban siswa. Dengan menghitung frekuensi kata-kata yang relevan, term frequency memberikan indikasi seberapa baik siswa memahami dan menyampaikan konsep yang diminta dalam pertanyaan ujian. Dengan kata lain, term frequency membantu mengevaluasi seberapa kaya konten jawaban siswa dan sejauh mana mereka menggambarkan pemahaman yang mendalam terhadap materi pelajaran.

2. Permasalahan

Proses penilaian jawaban ujian esai dilakukan dengan cara membaca setiap kata atau kalimat dalam jawaban ujian siswa secara bertahap dan membandingkannya dengan kata kunci jawaban yang telah ditentukan oleh guru. Lalu nilai diberikan berdasarkan tingkat kemiripan jawaban siswa dengan kunci jawaban yang telah dibuat oleh guru. Namun guru yang cenderung memeriksa jawaban ujian siswa satu per satu dengan menguraikan dan menyesuaikan jawaban esai siswa yang jawaban nya beragam sehingga cenderung berbeda dengan kunci jawaban. Hal itu dapat mempengaruhi objektivitas dalam penilaian, pengaruh bias ini dapat menyebabkan penilaian yang tidak konsisten antara satu siswa dengan siswa lainnya. Lalu untuk proses penilaian jawaban esai juga cenderung menggunakan waktu yang tidak relatif singkat jika dibandingkan dengan penilaian soal pilihan ganda, yang dimana perlu memeriksa jawaban siswa perkalimat.

Terdapat beberapa indikator yang memunculkan adanya suatu masalah dari segi guru yang diambil dari sampel guru sebanyak 16 guru mata pelajaran. Untuk permasalahan nya dijabarkan pada tabel yang berisi tiga pertanyaan kuesioner sebagaimana dijelaskan pada Tabel 1.

Tabel 1 Hasil Kuesioner Permasalahan Untuk Guru

No	Pertanyaan Kuesioner	Iya	Tidak
1	Apakah bapak/ibu guru terdapat kesulitan yang dialami dalam melakukan penilaian jawaban ujian esai?	75%	25%
2	Apakah bapak/ibu guru dalam melakukan penilaian jawaban ujian esai, memerlukan waktu yang relatif lama?	81,3%	18,8%
3	Apakah bapak/ibu guru pernah menerima keluhan dari siswa terkait hasil nilai yang diberikan kepada siswa?	31,3%	68,8%

Terdapat beberapa indikator yang memunculkan adanya suatu masalah dari segi siswa yang diambil dari sampel siswa sebanyak 24 siswa. Untuk permasalahan nya dijabarkan pada tabel yang berisi tiga pertanyaan kuesioner sebagaimana dijelaskan pada Tabel 2.

Tabel 2 Hasil Kuesioner Permasalahan Siswa

No	Pertanyaan Kuesioner	Iya	Tidak
1	Apakah Anda merasa ada siswa lain yang menerima skor lebih tinggi meskipun kualitas jawaban mereka serupa atau lebih rendah?	62,5%	37,5%
2	Apakah Anda merasa bahwa hasil penilaian esai sering diterima setelah waktu yang cukup lama sejak ujian berlangsung?	75%	25%

Dari uraian permasalahan di atas, dapat diketahui bahwa proses penilaian esai cenderung sulit karena harus menyesuaikan jawaban siswa yang memiliki kosakata dan cara penyampaian yang beragam, siswa merasa ada ketidaksesuaian atau ketidakadilan dalam pemberian skor, di mana kualitas jawaban serupa mendapatkan nilai yang berbeda dan proses penilaian esai membutuhkan waktu yang relatif lama dibandingkan dengan penilaian soal pilihan ganda.

3. Tujuan

Menerapkan algoritma *cosine similarity* dan *term frequency* untuk mendapatkan tingkat ketepatan, efektivitas, mengembangkan *prototype* aplikasi, dan mengukur seberapa tepat dan efektif algoritma yang diterapkan untuk penilaian esai.

B. LANDASAN TEORI

1. Penerapan Term Frequency

Istilah dalam dokumen diberi bobot berdasarkan frekuensinya dalam dokumen tersebut, metode paling sederhana adalah menggunakan frekuensi istilah (TF), disebut sebagai tft,d (frekuensi istilah t dalam dokumen d), bobot ini sama dengan jumlah kemunculan istilah dalam dokumen (Manning, Raghavan, & Schütze, 2009, hal. 117-118). dalam konteks suatu dokumen, set bobot yang ditentukan oleh TF, atau fungsi pembobotan serupa, berfungsi sebagai representasi kuantitatif dokumen dalam model bag of words, dalam model ini, urutan istilah yang tepat diabaikan, dan hanya jumlah kemunculan setiap istilah yang penting. Untuk rumus pembobotan TF dapat dijelaskan sebagai berikut (Manning, Raghavan, & Schütze, 2009, hal. 117-118):

$$TF(t, d) = \frac{\text{Jumlah kemunculan kata } t \text{ dalam dokumen } d}{\text{Total jumlah kata dalam dokumen } d}$$

Keterangan:

t : adalah kata atau term yang dihitung frekuensi nya
 d : adalah dokumen yang sedang di evaluasi

Dalam kata-kata, *Term frequency* (TF) untuk suatu term t dalam suatu dokumen d dihitung dengan membagi jumlah kemunculan term tersebut dalam dokumen dengan total jumlah term dalam dokumen.

Contoh kasus Misalkan ada dokumen berikut: Dokumen d : "kucing suka ikan, kucing bermain dengan kucing lain. Untuk Langkah-Langkah Menghitung Term Frequency (TF): Identifikasi setiap term unik dalam dokumen: Term yang unik dalam dokumen d adalah: "kucing", "suka", "ikan", "bermain", "dengan", "lain". Kemudian hitung jumlah kemunculan setiap term dalam dokumen: "kucing": 3 kali, "suka": 1 kali, "ikan": 1 kali, "bermain": 1 kali, "dengan": 1 kali, "lain": 1 kali. Setelah itu dihitung menggunakan rumus term frequency hasilnya seperti berikut ini:

Dokumen = {"kucing": 0.375, "suka": 0.125, "ikan": 0.125, "bermain": 0.125, "dengan": 0.125, "lain": 0.125}

2. Penerapan Cosine Similarity

Cosine similarity yaitu metode untuk mengukur kesamaan antara dua dokumen, dua dokumen tersebut masing-masing didefinisikan sebagai A , dan B , Dokumen dapat direpresentasikan sebagai vektor dalam ruang vektor, Tingkat kesamaan antara dua dokumen diukur berdasarkan besar sudut yang menghubungkan dua dokumen tersebut (Manning, Raghavan, & Schütze, 2009, hal. 120–121). Pada cosine similarity untuk tingkat kesamaan antara dua dokumen itu diukur berdasarkan besar sudut yang menghubungkan keduanya, semakin kecil sudutnya, maka semakin tinggi juga tingkat kesamaan antara kedua dokumen tersebut. Untuk rumus *Cosine similarity* yaitu sebagai berikut ini (Manning, Raghavan, & Schütze, 2009, hal. 120-121):

$$Sim(d_1, d_2) = \frac{\vec{V}(d_1) \cdot \vec{V}(d_2)}{|\vec{V}(d_1)| |\vec{V}(d_2)|}$$

Keterangan:

θ = sudut antara dua vektor,

$A \cdot B$ = dot product pada vektor A dan B dan dikalkulasikan sebagai

$$A \cdot B = A^T B = \sum_{i=1}^n A_i B_i = A_1 B_1 + A_2 B_2 + A_3 B_3 + \dots + A_n B_n,$$

$\|A\|$ dan $\|B\|$ = skala pada vektor yang dikalkulasikan sebagai $\|A\| = \sqrt{A_1^2 + A_2^2 + \dots + A_n^2}$ dan begitupun dengan $\|B\|$

Contoh penerapan Cosine Similarity dengan dua dokumen. Dokumen A : "kucing suka bermain dengan bola" dan Dokumen B : "kucing bermain dengan teman-temannya". Pertama representasikan setiap dokumen sebagai vektor berdasarkan dokumen yaitu Dokumen A : "kucing": 1, "suka": 1, "bermain": 1, "dengan": 1, "bola": 1,

Total kata dalam A = 5. Vektor A = (1/5, 1/5, 1/5, 1/5, 1/5). Dan Dokumen B: "kucing": 1, "bermain": 1 "dengan":1 "teman-temannya": 1, Total kata dalam B = 4. Vektor B = (1/4, 1/4, 1/4). Dot product A.B dihitung dengan mengalikan nilai yang sesuai antara kedua vektor dan menjumlahkan hasilnya.

$$A \cdot B = \left(\frac{1}{5} \times \frac{1}{4}\right) + \left(\frac{1}{5} \times 0\right) + \left(\frac{1}{5} \times \frac{1}{4}\right) + \left(\frac{1}{5} \times \frac{1}{4}\right)$$

$$A \cdot B = \frac{1}{20} + 0 + \frac{1}{20} + \frac{1}{20} = \frac{3}{20}$$

Selanjutnya hitung panjang (norm) dari masing-masing vektor:

$$\|A\| = \sqrt{\left(\frac{1}{5}\right)^2 + \left(\frac{1}{5}\right)^2 + \left(\frac{1}{5}\right)^2 + \left(\frac{1}{5}\right)^2 + \left(\frac{1}{5}\right)^2} = \sqrt{\frac{5}{25}} = \frac{1}{\sqrt{5}}$$

$$\|B\| = \sqrt{\left(\frac{1}{4}\right)^2 + (0)^2 + \left(\frac{1}{4}\right)^2 + \left(\frac{1}{4}\right)^2} = \sqrt{\frac{3}{16}} = \frac{\sqrt{3}}{4}$$

Sekarang kita dapat menghitung cosine similarity menggunakan rumus:

$$Sim(A,B) = \frac{\vec{V}(d_1) \cdot \vec{V}(d_2)}{|\vec{V}(d_1)| |\vec{V}(d_2)|}$$

$$Sim(A,B) = \frac{\frac{3}{20}}{\left|\frac{1}{\sqrt{5}}\right| \left|\frac{\sqrt{3}}{4}\right|} = \frac{12\sqrt{5}}{20\sqrt{3}} = 0.69$$

Maka untuk hasil tingkat kemiripan nya yaitu sebesar **0.69**

3. Penilaian Ujian Esai

Menjelaskan peningkatan kualitas pendidikan dapat dilihat dari nilai-nilai yang diperoleh peserta didik penilaian merupakan langkah lanjutan setelah dilakukan pengukuran. Menyatakan tes jenis uraian atau esai adalah jenis tes yang menuntut peserta didik untuk mengorganisasikan, mengekspresikan, serta menyatakan jawabannya menurut kata – kata sendiri, Tes esai dapat dibagi menjadi dua bentuk berdasarkan keluasan materi yang ditanyakan, yakni uraian terbatas dan uraian bebas (Widiyanto, 2018, hal. 124-125). Bentuk esai menuntut peserta untuk mengorganisir, menginterpretasi, dan menghubungkan konsep-konsep yang telah dipelajari, Ini merupakan ujian bagi peserta untuk menguji pemahaman mereka terhadap materi yang telah diberikan, serta kemampuan mereka untuk mengingatnya.

C. METODE

1. Metode Penelitian

Penelitian ini merupakan jenis penelitian dan pengembangan atau Research and Development (R&D). Metode penelitian Research and Development (R&D) adalah metode penelitian yang digunakan untuk mengembangkan suatu produk dan menguji keefektifan produk tersebut (Sugiyono, 2013, hal. 297).

2. Objek, Definisi Operasional Variabel dan Lokasi Penelitian

Penelitian ini berfokus pada proses penilaian jawaban esai siswa berdasarkan tingkat kemiripan dan kesesuaian dengan kunci jawaban. Penelitian dilakukan dalam konteks evaluasi pembelajaran di Sekolah Menengah Kejuruan (SMK) yang menggunakan soal esai sebagai metode penilaian pemahaman siswa. Tantangan utama dalam penilaian jawaban esai meliputi efisiensi waktu, konsistensi dalam pemberian nilai, serta potensi kesalahan interpretasi akibat variasi penyampaian jawaban siswa. Oleh karena itu, penelitian ini mengeksplorasi dan mengembangkan pendekatan berbasis teknologi yang dapat membantu guru dalam menilai jawaban esai secara lebih objektif, akurat, dan efisien.

Variabel dalam penelitian ini adalah keakuratan penilaian jawaban esai siswa sebagai variabel dependen, sedangkan variabel independennya adalah tingkat kemiripan jawaban, kesesuaian jawaban, efisiensi penilaian, dan konsistensi penilaian.

- Tingkat kemiripan jawaban diukur berdasarkan persentase kesamaan teks antara jawaban siswa dengan kunci jawaban menggunakan metode analisis teks berbasis algoritma.
- Kesesuaian jawaban diukur berdasarkan relevansi isi jawaban siswa dengan konsep yang terdapat dalam kunci jawaban, meskipun dengan perbedaan struktur kalimat atau pemilihan kata.
- Efisiensi penilaian diukur berdasarkan waktu yang dibutuhkan untuk menyelesaikan penilaian jawaban esai dalam skala waktu tertentu.
- Konsistensi penilaian diukur dengan membandingkan hasil penilaian sistem dengan penilaian manual oleh guru pada beberapa sampel jawaban esai.

Penelitian ini dilakukan di salah satu Sekolah Menengah Kejuruan (SMK) kota bogor yang menggunakan metode evaluasi berbasis esai dalam berbagai mata pelajaran, baik teori maupun praktik. Pemilihan lokasi penelitian didasarkan pada kebutuhan akan sistem penilaian yang lebih efisien dan objektif dalam menilai jawaban esai siswa.

3. Populasi dan Sampel

Populasi dalam penelitian ini terdiri dari guru dan siswa yang terlibat dalam proses evaluasi jawaban esai, serta dua orang ahli di bidang kecerdasan buatan yang bertindak sebagai evaluator dalam pengujian keakuratan sistem. Sampel dalam penelitian ini ditentukan menggunakan rumus Slovin dengan margin kesalahan sebesar 0,2. Sampel terdiri dari:

- Responden ahli, yaitu dua dosen yang memiliki keahlian dalam bidang kecerdasan buatan, yaitu Julio Warmansyah, S.Kom, M.MSI., dan Binanda Wicaksana, M.Kom. Pengujian terhadap ahli dilakukan menggunakan metode black box testing untuk mengevaluasi efektivitas sistem.
- Responden guru, dengan jumlah populasi 45 orang. Berdasarkan perhitungan rumus Slovin, jumlah sampel yang digunakan adalah 16 guru.
- Responden siswa, dengan jumlah populasi 379 orang. Berdasarkan perhitungan rumus Slovin, jumlah sampel yang digunakan adalah 24 siswa.
- Pengujian kepada responden guru dan siswa dilakukan menggunakan metode Post-Study System Usability Questionnaire (PSSUQ) untuk mengukur tingkat kepuasan dan efektivitas sistem dalam mendukung proses penilaian jawaban esai.

4. Teknik Pengumpulan Data

Data dalam penelitian ini dikumpulkan melalui beberapa metode, yaitu survei langsung, wawancara, observasi, dan eksperimen. Teknik pengumpulan data dibagi menjadi dua kategori berdasarkan jenis responden, yaitu ahli dan pengguna (guru dan siswa).

- Pengumpulan Data dari Ahli

Data yang dikumpulkan dari ahli bertujuan untuk mengevaluasi keakuratan dan efektivitas sistem dari perspektif teknis. Proses pengumpulan data dilakukan melalui metode black box testing, di mana dua dosen ahli dalam bidang kecerdasan buatan mengevaluasi fungsionalitas aplikasi yang telah dikembangkan.

- Pengumpulan Data dari Pengguna

Data yang dikumpulkan dari pengguna (guru dan siswa) bertujuan untuk mengevaluasi tingkat kepuasan dan kemudahan penggunaan sistem dalam mendukung proses penilaian jawaban esai. Pengumpulan data dilakukan menggunakan metode kuesioner Post-Study System Usability Questionnaire (PSSUQ), yang terdiri dari empat bagian utama, yaitu Overall, System Usefulness, Information Quality dan Interface Quality

5. Teknik Analisis Data

Analisis data dalam penelitian ini dilakukan menggunakan dua pendekatan utama, yaitu **metode persentase kelayakan** dan **evaluasi performa model klasifikasi menggunakan F1-score**.

- Metode Persentase Kelayakan**

Metode ini digunakan untuk mengevaluasi tingkat kelayakan sistem berdasarkan hasil yang diperoleh dari responden ahli dan pengguna. Persentase kelayakan dihitung menggunakan rumus berikut:

$$\text{persentase kelayakan (\%)} = \frac{\text{Skor yang didapatkan}}{\text{Skor Maksimal}} \times 100\%$$

Hasil perhitungan persentase ini digunakan untuk mengkategorikan tingkat kelayakan sistem berdasarkan tabel berikut:

Tabel 3 Interpretasi Persentase Kelayakan

Persentase Pencapaian	Interpretasi
<21%	Sangat Tidak Layak
21% - 40%	Tidak Layak
41% - 60%	Cukup Layak
61% - 75%	Layak
76% - 100%	Sangat Layak

Sumber: (Lubis dkk., 2023, hlm. 85)

Metode ini diterapkan untuk mengevaluasi hasil pengujian dari responden pengguna yang menilai aspek kelayakan sistem berdasarkan kuesioner PSSUQ yang telah disebarkan.

b. **Evaluasi Performa Model dengan F1-score**

Untuk menilai keakuratan model dalam menilai jawaban esai, digunakan metrik **F1-score** yang mempertimbangkan keseimbangan antara **precision** dan **recall**, terutama pada dataset yang memiliki distribusi kelas tidak seimbang.

F1-score dihitung menggunakan rumus berikut:

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Dengan perhitungan:

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

Keterangan:

- 1) TP (True Positive): Hasil yang dinilai benar oleh algoritma dan sesuai dengan nilai sistem yang diberikan;
- 2) FP (False Positive): Hasil yang dinilai benar oleh algoritma tetapi sebenarnya salah menurut nilai sistem;
- 3) FN (False Negative): Hasil yang dinilai salah oleh algoritma tetapi sebenarnya benar menurut nilai sistem;

Hasil F1-score diinterpretasikan berdasarkan tabel berikut:

Tabel 4 Interpretasi F1-Score

Nilai F1-Score	Interpretasi
≥ 0.90	Sangat Baik
0.75 – 0.89	Baik
0.50 – 0.74	Cukup Baik
< 0.50	Tidak Baik

Sumber: (Widodo, 2022, p. 24)

Evaluasi ini dilakukan untuk menilai sejauh mana sistem dapat memberikan hasil penilaian yang tepat dibandingkan dengan penilaian manual oleh guru.

D. HASIL DAN PEMBAHASAN

1. Hasil

Soal esai yang digunakan bersifat closed-ended, yaitu soal yang mengharuskan siswa menjawab berdasarkan teori. Soal ini membutuhkan jawaban yang relevan dan terstruktur sesuai dengan teori yang ada. Jawaban didapatkan dari soal esai yang telah dibuat sebelumnya beserta bobotnya.

Tabel 5 Soal dan Bobot

Soal	Bobot
Tuliskan pengertian proses fotosintesis yang terjadi pada tumbuhan!	20

Kunci jawaban dan jawaban siswa yang digunakan pada penelitian ini hanya berbentuk teks dan bukan berbentuk gambar, grafik, dan file. Untuk kunci jawaban dari data soal guru dan jawaban siswa diambil dari test ujian kepada siswa.

Tabel 6 Kunci Jawaban dan Jawaban Siswa

Kunci Jawaban	Jawaban Siswa
Fotosintesis adalah proses di mana tumbuhan menggunakan cahaya matahari, air, dan karbon dioksida untuk menghasilkan glukosa dan oksigen.	Fotosintesis yaitu proses tumbuhan yang menggunakan cahaya matahari dan karbon dioksida untuk menghasilkan oksigen.

Menentukan *Text-Preprocessing* dan Tokenisasi dari kunci jawaban dan jawaban siswa. Jawaban siswa dan kunci jawaban diproses untuk menghilangkan noise, seperti tanda baca, huruf kapital, dan karakter non-alfabet. Proses ini juga mencakup stemming untuk mengubah kata ke bentuk dasarnya.

Tabel 7 Hasil Stemming

Kunci Jawaban	Jawaban Siswa
fotosintesis proses tumbuhan menggunakan cahaya matahari air karbon dioksida menghasilkan glukosa oksigen	fotosintesis proses tumbuhan menggunakan cahaya matahari karbon dioksida menghasilkan oksigen

Setelah melakukan *text-preprocessing* selanjutnya dilakukan tokenisasi dengan cara pembersihan teks dari karakter non-alfanumerik dan memotong kalimat yang telah dibersihkan dipecah menjadi token (kata-kata individual) berdasarkan spasi.

Tabel 8 Hasil Text-Preprocessing dan Tokenisasi

Kunci Jawaban	Jawaban Siswa
["fotosintesis", "proses", "tumbuhan", "menggunakan", "cahaya", "matahari", "air", "karbon", "dioksida", "menghasilkan", "glukosa", "oksigen"]	["fotosintesis", "proses", "tumbuhan", "menggunakan", "cahaya", "matahari", "karbon", "dioksida", "menghasilkan", "oksigen"]

Selanjutnya kunci jawaban dan jawaban siswa akan dihitung frekuensi kemunculan nya per kata. Hasil dari tokenisasi di hitung dengan term frequency untuk menghitung frekuensi kata nya.

Tabel 9 Hasil Perhitungan Term Frequency

Kata	TF (Jawaban)	TF (Kunci)	DF	IDF	TF-IDF (Jawaban)	TF-IDF (Kunci)
fotosintesis	1	1	2	$\log(2/2+1)=0.0$	0.0	0.0
proses	1	1	2	$\log(2/2+1)=0.0$	0.0	0.0
tumbuhan	1	1	2	$\log(2/2+1)=0.0$	0.0	0.0
menggunakan	1	1	2	$\log(2/2+1)=0.0$	0.0	0.0
cahaya	1	1	2	$\log(2/2+1)=0.0$	0.0	0.0
matahari	1	1	2	$\log(2/2+1)=0.0$	0.0	0.0
air	0	1	1	$\log(2/1+1)=0.693$	0.0	0.693

Kata	TF (Jawaban)	TF (Kunci)	DF	IDF	TF-IDF (Jawaban)	TF-IDF (Kunci)
karbon	1	1	2	$\log(2/2+1)=0.0$	0.0	0.0
dioksida	1	1	2	$\log(2/2+1)=0.0$	0.0	0.0
menghasilkan	1	1	2	$\log(2/2+1)=0.0$	0.0	0.0
oksigen	1	1	2	$\log(2/2+1)=0.0$	0.0	0.0
glukosa	0	1	1	$\log(2/1+1)=0.693$	0.0	0.693

Setelah menghitung term frequency nya selanjutnya akan dihitung dengan cosine similarity untuk di ukur tingkat kemiripan nya. Pengukuran nya di hitung berdasarkan vektor frekuensi dari dua dokumen yaitu kunci jawaban dan jawaban siswa.

Perhitungan:

1) Dot Product:

$$2) = (1 \times 1) + (1 \times 1) + (1 \times 1) + (1 \times 1) + (1 \times 1) + (1 \times 1) + (0 \times 1) + (1 \times 1) + (1 \times 1) + (1 \times 1) + (1 \times 1) + (0 \times 1) = 10$$

3) Magnitude Jawaban Siswa:

$$\sqrt{1^2 + 1^2 + 1^2 + 1^2 + 1^2 + 1^2 + 0^2 + 1^2 + 1^2 + 1^2 + 1^2 + 0^2} = \sqrt{10} = 3.162$$

4) Magnitude Kunci Jawaban:

$$\sqrt{1^2 + 1^2 + 1^2 + 1^2 + 1^2 + 1^2 + 1^2 + 1^2 + 1^2 + 1^2 + 1^2 + 1^2} = \sqrt{12} = 3.464$$

5) Cosine Similarity:

$$= \frac{10}{3.162 \times 3.464} = \frac{10}{10.953} = 0.912 \text{ atau } 91\%$$

Setelah hasil dari perhitungan term frequency dan cosine similarity maka hasil nilai nya akan di tentukan masuk kedalam grade yang mana, untuk grade nya yaitu:

if final score = 100,

Sempurna! Jawaban siswa sama dengan kunci jawaban guru.

if final score ≥ 90,

"Sangat baik! Jawaban siswa sangat mirip dengan kunci jawaban guru."

if final score ≥ 80,

"Baik. Jawaban siswa cukup mirip dengan kunci jawaban guru."

f final score ≥ 70,

"Cukup. Jawaban siswa memiliki beberapa kesamaan dengan kunci jawaban guru."

if final score ≥ 60,

"Kurang. Jawaban siswa memiliki sedikit kesamaan dengan kunci jawaban guru."

else,

"Sangat kurang. Jawaban siswa tidak mirip dengan kunci jawaban guru."

Maka grade dari nilai siswa itu adalah "Sangat baik! Jawaban siswa sangat mirip dengan kunci jawaban guru."

2. Pembahasan

Uji hasil yang dilakukan pada penelitian ini dengan cara membandingkan nilai hasil dari pengerjaan ujian esai siswa (nilai aktual) dengan nilai hasil penilaian dari algoritma (nilai algoritma), untuk uji hasil ini menggunakan metode F1-Score. Nilai aktual diperoleh dari hasil pemeriksaan langsung oleh guru berdasarkan pedoman penilaian yang telah ditetapkan. Hasil pemeriksaan guru (nilai aktual) dibandingkan dengan hasil

penilaian algoritma (nilai algoritma) untuk mengukur ketepatan algoritma. Proses perbandingan dilakukan menggunakan metode evaluasi seperti F1-score, dengan membandingkan hasil benar dan salah dari algoritma terhadap hasil aktual. Data hasil nilai siswa dari pemeriksaan guru dan algoritma dapat dilihat pada tabel.

Tabel 10 Hasil Nilai Ujian Esai

No	Nama	Nilai Aktual	Nilai Algoritma	TP	FP	FN
1	Muhammad Rasya S.	75	76	1	0	0
2	Raya Aulia Zahrani	75	71	0	0	1
3	Pandu Bhekti Febrian	80	84	0	1	0
4	Muhammad Riva Ramdhan	70	71	1	0	0
5	Pandu Dwi Pangestu	85	84	1	0	0
6	Sabil Nadir Air Raya	75	76	1	0	0
7	Nadisa Putri Sabina	80	84	0	1	0
8	Syahbian Rizqi Ramadan	85	88	1	0	0
9	Muhammad Feriansyah	85	88	1	0	0
10	Cristen Nathanael	70	71	1	0	0

Penjelasan TP, FP dan FN:

- TP (True Positive): Hasil yang didapatkan oleh algoritma sesuai atau mendekati dari nilai yang sebenarnya diberikan oleh guru.
- FP (False Positive): Hasil didapatkan oleh algoritma lebih tinggi dari nilai yang sebenarnya diberikan oleh guru.
- FN (False Negative): Hasil didapatkan oleh algoritma lebih rendah dari nilai yang sebenarnya diberikan oleh guru.

Perhitungan Precision, Recall, dan F1-Score:

- Precision :

$$Precision = \frac{TP}{TP + FP} = \frac{7}{7 + 2} = \frac{7}{9} = 0.778$$

- Recall :

$$Recall = \frac{TP}{TP + FN} = \frac{7}{7 + 1} = \frac{7}{8} = 0.875$$

- F1-Score :

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} = 2 \times \frac{0.778 \times 0.875}{0.778 + 0.875} = 2 \times \frac{0.682}{1.653} = 0.824$$

Nilai F1-score yang didapatkan dari perhitungan di atas adalah sebesar 0.824 masuk dalam kategori **Baik**. Hal ini menunjukkan bahwa algoritma cosine similarity dan term frequency memiliki performa yang cukup tepat dalam menilai jawaban esai siswa dibandingkan dengan hasil penilaian guru.

E. KESIMPULAN

Berdasarkan hasil penelitian yang dilakukan, dapat diuraikan beberapa kesimpulan, antara lain:

- Mendapatkan hasil nilai perolehan jawaban esai secara tepat dengan menggunakan algoritma Term Frequency dan Cosine Similarity.
- Prototype* aplikasi penilaian esai berhasil dikembangkan untuk mempermudah guru dalam proses pemberian nilai esai.
- Untuk uji hasil dengan menggunakan F1-Score mendapatkan hasil sebesar 0.824 yang menunjukkan bahwa sistem memiliki ketepatan yang masuk dalam kategori "Baik".
- Untuk uji ahli sistem ini mendapatkan nilai 100% untuk kelayakan nya. Lalu untuk uji pengguna nya mendapatkan hasil secara keseluruhan sebesar 86.63%, kegunaan sistem sebesar 87.67%, kualitas informasi sebesar 87.25%, dan kualitas antarmuka sebesar 80.79% yang artinya aplikasi ini "Sangat Layak" digunakan.

F. DAFTAR PUSTAKA

- [1] Hadid, Marwan, dan Triawan, Anggra. (2025). Penerapan Algoritma Winnowing dan Cosine Similarity untuk Penilaian Jawaban Esai. *Jurnal Ilmiah Saintekom*, Volume 01 No. 01. Juni 2025; 66 – 72.
- [2] Heaton, J. (2013). *Artificial Intelligence for Humans, Volume 1: Fundamental Algorithms*. Washington: Heaton Research, Inc.
- [3] Manning, C. D., Raghavan, P., & Schütze, H. (2009). *An Introduction to Information Retrival*. Cambridge: Cambridge University Press.
- [4] Roger, P. S., & Maxim, B. R. (2015). *Software Engineering: A Practitioner's Approach* (8 ed.). New York: McGraw-Hill Education.
- [5] Russel, S. J., & Norvig, P. (2010). *Artificial Intelligence A Modern Approach* (3 ed.). Upper Saddle River: Prentice Hall.
- [6] S., R. A., & Shalahuddin, M. (2016). *Rekayasa Perangkat Lunak (Terstruktur dan Berorientasi Objek)*. Bandung: Informatika.
- [7] Smolentzov, A. (2012). *Automated Essay Scoring for Swedish*. Stockholm: Stockholm University.
- [8] Sommerville, I. (2016). *Software Engineering* (10 Ed.). Pearson Education.
- [9] Sugiyono. (2013). *Metode Penelitian Kuantitatif, Kualitatif Dan R & D* (19 Ed.). Bandung: Alfabeta.
- [10] Sugiyono. (2016). *Metode Penelitian Kuantitatf, Kualitatif Dan R&D*. Bandung: Alfabeta.
- [11] Widiyanto, J. (2018). *Evaluasi Pembelajaran*. Madiun: Unipma Press.
- [12] Widodo, R. B. (2022). *Machine Learning Metode K-Nearest Neighbors Klasifikasi Angka Bahasa Isyarat*. Malang: Media Nusa Creative.
- [13] Zulfikar, R., Sari, F. P., Fatmayati, A., Wandini, K., Haryati, T., Jumini, S., Fadilah, H. (2024). *Metode Penelitian Kuantitatif Teori, Metode Dan Praktik*. Bandung: Widina Media Utama.