

## PENERAPAN METODE COSINE SIMILARITY UNTUK PENGECEKAN KEMIRIPAN JAWABAN UJIAN SISWA

Firmansyah Fataruba

Program Studi Teknik Informatika S1, Fakultas Teknologi Industri  
Institut Teknologi Nasional Malang, Jalan Raya Karanglo km 2 Malang, Indonesia  
*Firmansyah.fataruba@gmail.com*

### ABSTRAK

Ujian bertipe essay merupakan pilihan evaluasi yang di berikan guru di mana pilihan jawaban tidak di sediakan, dan siswa harus menjawab soal dengan menggunakan kalimat, ujian essay sangat membantu siswa dalam melatih kemampuan dari pelajaran guru yang di berikan kepada siswa. Ujian essay juga dapat membantu pemahaman pada siswa dan mengevaluasi tingkat kemampuan pada siswa itu sendiri. Ujian dengan sistem ini sampai sekarang masih sangat sering digunakan karena lebih dapat mengukur tingkat pencapaian pembelajaran siswa dengan objektif karena siswa di tuntut untuk mengisi dengan kemampuan kalimat bukan seperti halnya pilihan ganda. Pengajar yang memberikan ujian essay akan memerlukan waktu yang sangat lama untuk mengoreksi jawaban ujian essay siswa yang banyak dan itu dapat mempengaruhi kualitas penilaian terhadap nilai yang di berikan.

Tujuan dari penelitian ini adalah untuk membantu pengajar memberikan nilai yang objektif dengan menggunakan metode cosine similarity pada sistem agar dapat melakukan penilaian jawaban essay dengan membandingkan kunci jawaban pengajar dengan jawaban peserta didik. Dan juga dapat membandingkan tingkat kemiripan antara siswa satu dengan siswa yang lain. Cosine similarity adalah metode untuk menghitung kesamaan dan kemiripan dari dua dokumen, dokumen yang di hitung perbandingannya dalam penelitian ini adalah kunci jawaban pengajar dan jawaban siswa. Untuk menyamakan frekuensi setiap kata pada kalimat yang ada di dokumen di gunakan persamaan Tf atau Term Frequency , Term Frequency merupakan faktor yang menentukan bobot kata yang di dasarkan pada jumlah frekuensi kata dalam suatu dokumen.

Berdasarkan hasil penelitian bahwa sistem pengecekan kemiripan jawaban siswa menggunakan metode cosine similarity telah berjalan dengan baik untuk ujian essay Biologi Hasil uji coba menunjukkan kesesuaian nilai sistem dengan nilai yang diberikan oleh pengajar tingkat akurasinya 80% menggunakan Confusion Matrix.

**Kata kunci :** *Cosine Similarity, Ujian Essay, Pembobotan, term Frequency, codeigniter*

### 1. PENDAHULUAN

Kemajuan ilmu pengetahuan dan teknologi memiliki pengaruh yang sangat besar dalam berbagai bidang kehidupan manusia [1]. Pendidikan sebagai salah satu bagian yang tidak terpisahkan dari proses pendewasaan manusia tentu di satu sisi memiliki andil yang besar bagi pengembangan ilmu pengetahuan dan teknologi tersebut, namun di sisi lain pendidikan juga perlu memanfaatkan kemajuan ilmu pengetahuan dan teknologi agar mampu mencapai tujuannya secara efektif dan efisien.

Dengan sistem essay merupakan bentuk evaluasi dimana pilihan jawaban tidak disediakan, dan siswa harus menjawab dengan kalimat, sehingga dapat melatih siswa dalam menyampaikan sesuatu informasi secara verbal, selain itu ujian esai juga menuntut pemahaman yang lebih baik akan suatu ilmu dan dapat digunakan untuk mengukur tingkat pemahaman manusia akan suatu ilmu secara lebih mendalam [1]. Pengajar memerlukan waktu yang banyak untuk memeriksa jawaban essay, semakin banyak jumlah ujian dan banyaknya jumlah pelajar yang mengikuti ujian, maka semakin banyak jumlah ujian yang dikoreksi oleh pengajar. Hal ini menyebabkan kualitas penilaian menurun dan terkadang penilaian tidak bersifat objektif lagi. [1].

Dalam kenyataannya sebagian besar guru masih belum bisa memberikan nilai yang objektif, karena masih melakukan penilaian terhadap peserta didik secara konvensional dan faktor lain seperti melewati bacaan yang panjang karena harus memeriksa banyak hasil ujian essay dari setiap peserta didik. Pemberian nilai yang objektif sangat berpengaruh dengan prestasi siswa agar siswa dapat mengukur kemampuannya dalam memahami pelajaran.

Oleh sebab itu dari latar belakang masalah di atas, maka penulis ingin membuat aplikasi berbasis website untuk memberikan penilaian pada pekerjaan soal essay yang di lakukan oleh peserta didik secara objektif dengan menggunakan metode cosine similarity. Dan pada penelitian ini penulis melakukan ujicoba sistem tersebut pada SMA Negeri 11 Ambon, Kota Ambon, Provinsi Maluku.

### 2. TINJAUAN PUSTAKA

#### 2.1. Penelitian Terkait

Ujian dengan sistem esai merupakan bentuk evaluasi dimana pilihan jawaban tidak disediakan, dan siswa harus menjawab dengan kalimat, sehingga dapat melatih siswa dalam menyampaikan sesuatu informasi secara verbal, selain itu ujian esai juga

menuntut pemahaman yang lebih baik akan suatu ilmu dan dapat digunakan untuk mengukur tingkat pemahaman manusia akan suatu ilmu secara lebih mendalam. Ujian dengan system ini tetap menjadi pilihan pengajar untuk mengevaluasi tingkat kemampuan pemahaman siswa walaupun kenyataannya tidak mudah untuk memberikan penilaian yang objektif pada jawaban siswa. Pengajar memerlukan waktu yang banyak untuk memeriksa jawaban essay, semakin banyak jumlah ujian dan banyaknya jumlah pelajar yang mengikuti ujian, maka semakin banyak jumlah ujian yang dikoreksi oleh pengajar. Hal ini menyebabkan kualitas penilaian menurun dan terkadang penilaian tidak bersifat objektif lagi. Beragam metode yang dapat digunakan untuk melakukan penilaian jawaban ujian esai secara otomatis salah satunya yang digunakan adalah metode *cosine similarity*. Metode Cosine Similarity adalah metode untuk menghitung kesamaan atau kemiripan dari dua buah dokumen, pada penelitian ini digunakan dokumen jawaban ujian siswa dan dokumen kunci jawaban pengajar. Untuk menyamakan frekuensi setiap kata yang terdapat didalam kalimat yang ada digunakan persamaan Tf atau *Term Frequency*, *Term Frequency* merupakan suatu faktor yang menentukan bobot kata yang didasarkan pada jumlah frekuensi kata dalam sebuah dokumen. Berdasarkan hasil penelitian bahwa sistem penilaian esai otomatis menggunakan metode cosine similarity telah berjalan dengan baik untuk ujian esai dalam bahasa Inggris Hasil uji coba menunjukkan kesesuaian nilai sistem dengan nilai yang diberikan oleh pengajar adalah rata-rata 89,48% [1]

Penilaian hasil ujian secara manual terkadang memiliki kelemahan yaitu terdapat penilaian yang berbeda terhadap peserta-peserta ajar dengan jawaban yang hampir sama. Pemberian nilai otomatis dilakukan dengan menggunakan stemming Porter yang dikombinasikan dengan *Cosine similarity*. Algoritma Porter digunakan untuk mencari kata dasar dari kata-kata yang digunakan dalam kunci jawaban. Hasil dari ekstraksi kata dasar pada kunci jawaban dan jawaban siswa akan dibandingkan dengan menggunakan *Cosine similarity*. [2]

Pemberian nilai otomatis dilakukan dengan menggunakan stemming Porter yang dikombinasikan dengan *Cosine similarity*. Hasil dari ekstraksi kata dasar pada kunci jawaban dan jawaban siswa akan dibandingkan dengan menggunakan *Cosine similarity*. Proses pemodelan dilakukan dengan mengembangkan dari model yang telah ada dengan menyesuaikan konsep realtime yang akan direncanakan. Model yang didapatkan adalah proses text stemming dan proses ekstrak kata dari jawaban dilakukan pada komputer client (pengajar dan siswa) dan tidak membebankan ke sever. [2]

## 2.2. Preprocessing data Text

Preprocessing adalah tahap proses awal text mining terhadap teks untuk mempersiapkan teks menjadi data yang dapat diolah lebih lanjut. [4]

### 1. Tokenizing

Tahap *tokenizing* atau parsing adalah tahap pemotongan string input berdasarkan tiap kata yang menyusunnya. Pada prinsipnya proses ini adalah memisahkan setiap kata yang menyusun suatu dokumen. [5]

### 2. Case Folding

Case folding, merupakan proses mengubah semua huruf dalam dokumen menjadi huruf kecil. Hanya huruf ‘a’ sampai dengan ‘z’ yang diterima karakter-karakter dan tanda baca lainnya selain huruf juga dihilangkan. Tidak semua dokumen teks konsisten dalam penggunaan huruf capital. [6]

### 3. Filtering

Tahap *filtering* adalah tahap mengambil kata-kata penting dari hasil token. Pada tahap ini akan dilakukan pembersihan *tweet* dari spesial karakter, URL link, *username*, serta emoticon. [5]

### 4. Stemming

*Stemming* adalah suatu teknik pencarian bentuk dasar dari suatu *term*. Yang dimaksud dengan *term* itu sendiri adalah tiap kata yang berada pada suatu dokumen teks. [7]

## 2.3. Cosine Similarity

*Cosine Similarity* adalah ukuran kesamaan antara dua buah *vektor* dalam sebuah ruang dimensi yang didapat dari nilai cosinus sudut dari perkalian dua buah *vektor* yang dibandingkan karena *cosinus* dari 0° adalah 1 dan kurang dari 1 untuk nilai sudut yang lain, maka nilai *similarity* dari dua buah *vektor* dikatakan mirip ketika nilai dari *Cosine Similarity* adalah 1. *Cosine similarity* digunakan dalam ruang positif, dimana hasilnya dibatasi antara nilai 0 dan 1. Kalau nilainya 0 maka dokumen tersebut dikatakan mirip jika hasilnya 1 maka nilai tersebut dikatakan tidak mirip Perhatikan bahwa batas ini berlaku untuk sejumlah dimensi. Berikut adalah rumus *cosine similarity* Rumus umum *cosine similarity* [8]

$$\frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i x B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}} \quad (1)$$

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 vector

Dimana A merupakan bobot setiap ciri pada vector A, dan B merupakan bobot setiap ciri pada B. jika dikaitkan dengan *information retrieval* maka A adalah bobot setiap istilah pada dokumen A, dan B

merupakan bobot setiap istilah pada dokumen B. Pada penelitian ini digunakan *cosine similarity* karena citra merupakan salah satu data yang memiliki dimensi tinggi. Pada citra dapat dikatakan bahwa setiap *pixel* merupakan dimensi yang berbeda dan nilai warna pada setiap *pixel* tersebut merupakan nilai dari setiap dimensi. Pengukuran kemiripan dapat dilakukan dengan membandingkan dokumen 1 dengan dokumen 2 kemudian sistem akan menghitung nilai kemiripan.  $A_i B_i$  adalah nilai yang diperoleh dari term A dan term B kemudian kedua nilai tersebut dijumlahkan, kemudian nilai  $A_i$  semua nilai term dokumen A semua nilainya dipangkatkan dua, begitu juga dengan *term B\_i2* semua nilai yang diperoleh dipangkatkan dua kemudian semua nilai yang diperoleh dijumlahkan. [8]

**2.4. Pembobotan Term Frequency (Tf)**

Term Frequency adalah untuk menghitung kemunculan kata kemudian menyimpannya kedalam basis data tabel indeks. Selanjutnya adalah proses perhitungan kemiripan adalah proses untuk mencari kemiripan antar dokumen teks. Kemudian sistem menghitung nilai kemiripan dokumen A dan dokumen B maka user bisa melihat seberapa banyak dokumen yang mirip. [8]

Tabel 1. Contoh pembobotan hubungan antar kata [8]

Tid	Term	T (A)	T (B)
T1	Teknik	2	2
T2	Informatika	1	1
T3	Fakultas	1	0
T4	Universitas	1	0
T5	Sam	1	0
T6	Ratulangi	1	0
T7	Manado	1	0
T8	Program	0	1
T9	Studi	0	1
T10	Jurusan	0	1
T 11	Elektro	0	1

Vektor A dan vektor B tersebut mewakili setiap term “A” dan term “B” untuk melihat berapa banyak nilai yang diperoleh dalam setiap kata yang unik dalam dokumen, Setelah itu setiap term dalam vektor akan di hitung menggunakan Cosine Similarity. [8]

**2.5. Confusion Matrix**

*Confusion matrix* adalah sebuah tabel yang menyatakan jumlah data uji yang benar diklasifikasikan dan jumlah data uji yang salah diklasifikasikan. Contoh *confusion matrix*[11]

Table 2. Confusion Matrix untuk klasifikasi biner

		Kelas Prediksi	
		1	0
Kelas Sebenarnya	1	TP	FN
	0	FP	TN

Keterangan untuk Tabel 2 dinyatakan sebagai berikut:

- *True Positive* (TP), yaitu jumlah dokumen dari kelas 1 yang benar diklasifikasikan sebagai kelas 1.
- *True Negative* (TN), yaitu jumlah dokumen dari kelas 0 yang benar diklasifikasikan sebagai kelas 0.
- *False Positive* (FP), yaitu jumlah dokumen dari kelas 0 yang salah diklasifikasikan sebagai kelas 1.

*False Negative* (FN) yaitu jumlah dokumen dari kelas 1 yang salah diklasifikasikan sebagai kelas 0.

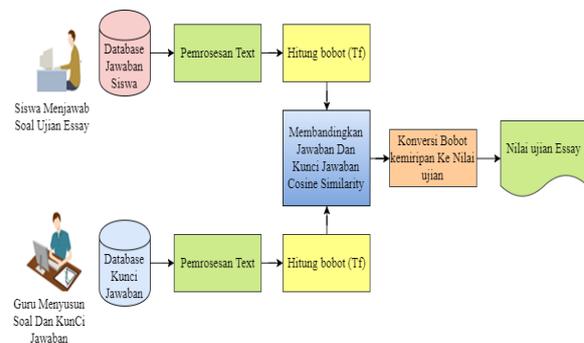
Perhitungan akurasi dinyatakan dalam persamaan (2)[11]

$$Akurasi = \frac{TP + TN}{TP + FN + FP + TN} \times 100\% (2)$$

**3. ANALISA DAN PERANCANGAN**

Penelitian pada penerapan metode *cosine similarity* menggunakan website untuk pengecekan kemiripan jawaban ujian siswa diharapkan dapat membantu pengajar memberikan nilai yang lebih objektif, dan menghemat waktu pengajar dalam memeriksa jawaban-jawaban dari peserta didik

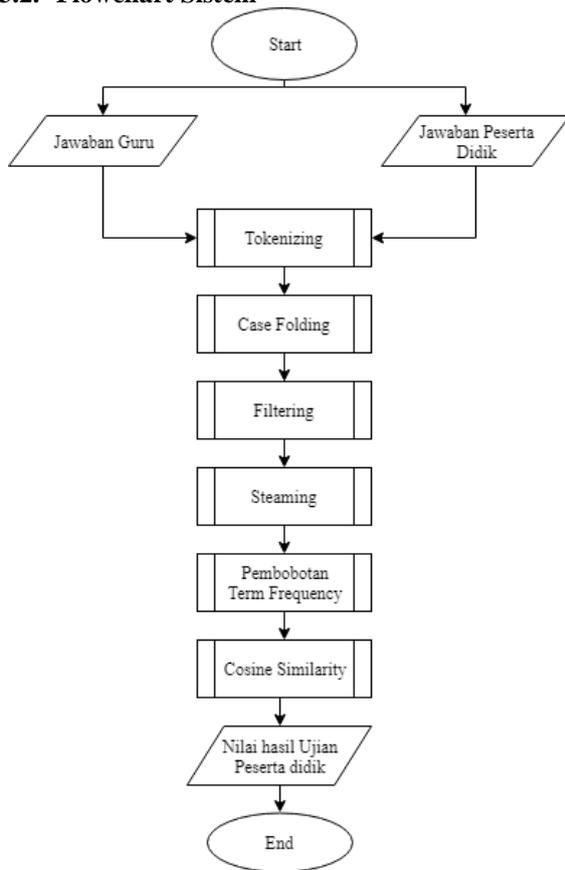
**3.1. Blok Diagram**



Gambar 1. Blok diagram system

Penjelasan dari Gambar 1 blok diagram sistem, Guru menyusun soal dan kunci jawaban setelah itu jawaban disimpan di *database* dan di pemrosesan teks (*Procesing data*) hitung bobot. Pada siswa akan di tampilkan soal dan kolom jawaban dan jawaban yang di inputkan siswa akan di proses dan di hitung bobot juga. Terakhir bobot dari jawaban siswa dan kunci jawaban akan di bandingkan tingkat kemiripannya menggunakan *Cosine Similarity* dan di konversi nilai kemiripan ke nilai ujian.

3.2. Flowchart Sistem



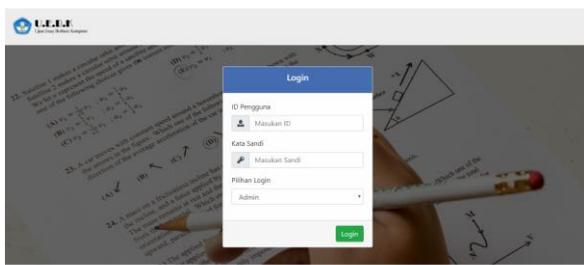
Gambar 2. Flowchart sistem

Pada Gambar 2 Flowchart sistem pada penelitian dimulai dengan sistem menerima input kunci jawaban dari guru dan siswa dan di inputkan ke database sistem, Setelah itu kunci jawaban dan jawaban siswa di preprosesing yang meliputi *Tokenizing*, *Filtering*, dan *Steaming*. Hasil dari kunci jawaban dan jawaban siswa setelah di *preprosesing*, akan di lakukan proses pembobotan (Tf) dan bobot yang di dapat akan di cek kemiripannya menggunakan *Cosine similarity*, Terakhir sistem akan memberikan nilai dari hasil klasifikasi.

4. IMPLEMENTASI DAN PENGUJIAN

4.1. Halaman login admin, guru dan siswa

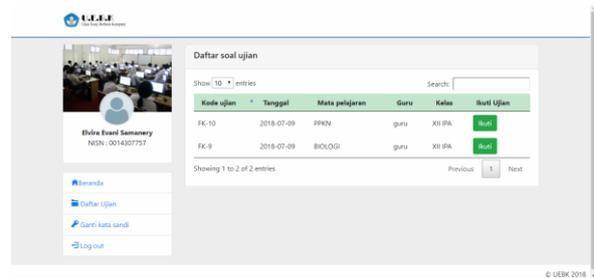
Tampilan login ini admin, guru dan siswa dapat login sebelum masuk ke halaman selanjutnya



Gambar 3. Halaman Login

4.2. Halaman daftar ujian essay siswa

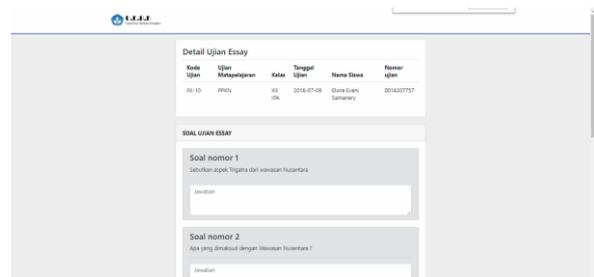
Halaman daftar ujian sebelum siswa mengikuti ujian, pada tampilan ini siswa akan memilih mata ujian yang akan dia ikuti dan memasukan token agar dapat mengikuti ujian yang ada pada daftar ujian



Gambar 4. Tampilan halaman ikuti ujian siswa

4.3. Halaman soal dan inputan jawaban siswa

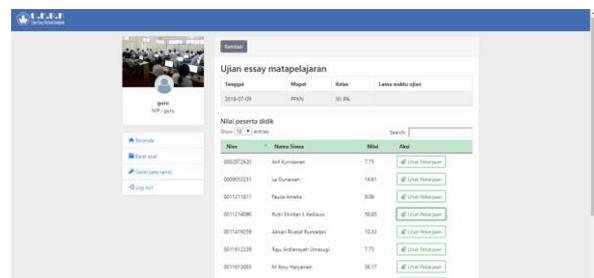
Tampilan halaman soal saat siswa sudah masuk ke ujian yang ingin di ikuti, dalam halaman ini siswa akan menginputkan jawabannya pada kolom jawaban yang telah di sediakan.



Gambar 5. Tampilan halaman soal dan inputan jawaban

4.4. Halaman monitoring nilai siswa

Pada halaman monitoring nilai siswa, guru dapat melihat hasil nilai dari siswa yang telah mengikuti ujian yang telah guru buat



Gambar 6. Tampilan monitoring nilai siswa

4.5. Halaman tingkat kemiripan dan penilaian sistem

Pada halaman ini guru dapat melihat tingkat kemiripan bobot dari setiap hasil jawaban siswa, nilai yang di peroleh dari tingkat kemiripan dan juga dapat melihat jawaban siswa yang telah di kerjakan.

Nomor Identitas Siswa : 0011214090  
 Nama Siswa : Putri Shintari L. Kelilaw

**Soal dan Jawaban Siswa**

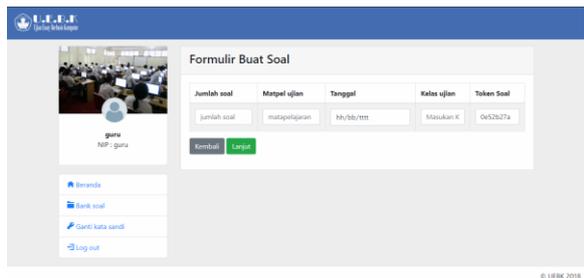
Soal ujian nomor 1 Sebutkan aspek Trigatra dari wawasan Nusantara Jawaban: letak dan bentuk geografi, keadaan dan kekayaan alam	Nilai benar 0.730%
Soal ujian nomor 2 Apa yang dimaksud dengan Wawasan Nusantara? Jawaban: Wawasan Nusantara adalah cara pandang dan sikap bangsa Indonesia mengenai diri dan bentuk geografinya berdasarkan Pancasila	Nilai benar 0.603%
Soal ujian nomor 3 Sebutkan aspek Pancasila dari wawasan Nusantara? Jawaban: 1. Ideologi 2. Politik 3. Ekonomi 4. Sosial Budaya 5. Pertahanan Keamanan	Nilai benar 0.655%
Soal ujian nomor 4 Apa arti semboyan Bhinneka tunggal ika? Jawaban: Artinya dari Bhinneka Tunggal Ika adalah Berbeda-beda tetapi tetap satu	Nilai benar 0.866%
Soal ujian nomor 5 Sebutkan Faktor Penghambat Integrasi Nasional Jawaban: - Kurangnya jitra sosial dalam masyarakat - Pembangunan yang tidak merata	Nilai benar 0.139%

**NILAI SISWA : 59.85**

Gambar 7. Tampilan tingkat bobot kemiripan jawaban siswa dan guru beserta nilai

**4.6. Halaman formulir untuk penginputan soal dan kunci jawaban guru**

Pada halaman ini guru pertama-tama menginputkan formulir soal, berikut formulir yang harus di isi sebelum menginputkan soal dan kunci jawaban adalah Jumlah soal, Matapelajaran, tanggal ujian, kelas ujian, dan kode token soal.



Gambar 8. Tampilan formulir menginputkan soal dan kunci jawaban

**4.7. Halaman penginputan soal dan kunci jawaban oleh guru**

Pada halaman ini guru akan menginputkan soal beserta kunci jawaban agar bisa di ujikan ke siswa.



Gambar 9. penginputan soal dan kunci jawaban

**4.8. Pengujian Metode Pada Sistem**

Pengujian perhitungan manual metode cosine similarity pada sistem adalah pengujian untuk mengetahui metode yang di terapkan di sistem dapat memberikan nilai tingkat kemiripan. Pengujian di lakukan dengan 1 buah soal dan kunci jawaban dan di bandingkan dengan 3 jawaban dari siswa.

Tabel 3. Hasil pengujian metode pada sistem

Siswa yang di uji	Bobot tingkat kemiripan kunci jawaban dan jawaban siswa yang metode berikan pada sistem
Selvira R Yusran	0.48
Akbar Arif Silawane	0.57
Poetri Fachnun Ayur veda Salatalohy	0.63

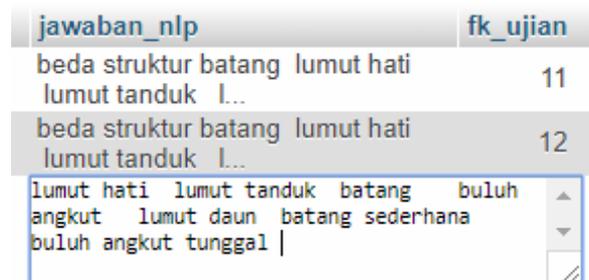
Berikut ini adalah contoh sampel perhitungan manual untuk mendapatkan bobot kemiripan dengan metode cosine similarity. Prosedur perhitungan yang di lakukan adalah :

1. Kunci jawaban dari guru akan dilakukan preprosesing. Soal dan Kunci Jawaban guru dan hasil kunci jawaban yang telah di preprosesing dapat di lihat pada Tabel 4

Tabel 4. Soal dan Kunci Jawaban dan hasil preprosesing kunci jawaban

Soal Guru	Kunci Jawaban Guru	Hasil Preprosesing
Jelaskan perbedaan struktur batang lumut hati, lumut tanduk, dan lumut daun!	<ul style="list-style-type: none"> <li>• Lumut hati dan lumut tanduk tidak berbatang dan tidak mempunyai pembuluh angkut.</li> <li>• Lumut daun mempunyai batang sederhana dengan pembuluh angkut tunggal.</li> </ul>	lumut hati lumut tanduk batang buluh angkut lumut daun batang sederhana buluh angkut tunggal

Hasil dari preprosesing ini sesuai dengan yang ada pada database sistem seperti di tunjukan pada Gambar



Gambar 10. Hasil Preprosesing kunci jawaban guru

2. Lakukan preprosesing pada jawaban siswa agar dapat di hitung bobot kemiripannya dengan kunci jawaban yang telah di preprosesing juga. Hasil dari preprosesing jawaban siswa

Tabel 5. Hasil preprosesing

Dokumen	Nama	Jawaban siswa	Hasil Jawaban yang telah di Preprosesing
1	Selvira R Yusran	lumut hati berstruktur lembek, lumut tanduk berstruktur kuat dan lumut daun strukturnya sering tumbuh dibagian luar	lumut hati struktur lembek lumut tanduk berstruktur kuat lumut daun struktur tumbuh
2	Akbar Arif Silawane	lumut hati berbentuk lembaran lumut tanduk berbentuk seperti kapsul memanjang seperti tanduk lumut daun memiliki batang	lumut hati bentuk lembaran lumut tanduk bentuk kapsul anjang tanduk lumut daun milik batang tegak cabang
3	Poetri Fachnun Ayur veda Salatalohy	lumut tanduk, batangnya lebih keras, berbentuk seperti lumut hati tapi sporofitnya berbentukkapsul memanjang, lumut hati, lumut dengan bentuk lembaran yang ditemukan didaerah lembap, lumut daun, lumut ini mempunyai bentuk batang yang tegak dan bercabang.	lumut tanduk batang, keras bentuk lumut hati sporofit berbentukkapsul memanjang lumut hati lumut bentuk lembaran temu daerah lembap lumut daun lumut bentuk batang tegak bercabang

3. Hasil preprosesing jawaban siswa sesuai dengan yang ada pada database sistem

id_jawaban	jawaban_siswa	jawaban_nlp	fk_soal	id_ujian	id_user
881	batang lumut hati, halus dan juga kecil, banyak ber...	batang lumut hati halus rada mkn batang lumu...	29	13	178
882	pada lumut hati, struktur batang tidak begitu koko...	lumut hati struktur batang kokoh lumut tanduk...	29	13	152
883	lumut hati, lumut dengan bentuk lembaran hidup yan...	lumut hati lumut bentuk lembaran hidup temu daer...	29	13	150
884	1. lumut hati tumbuh di daerah lembab 2. lumut tan...	1 lumut hati tumbuh daerah lembab 2. lumut tandu...	29	13	149
885	perbedaan struktur batang pada lumut hati, lumut t...	beda struktur batang lumut hati lumut tanduk lu...	29	13	157

Gambar 11. Hasil preprosesing jawaban siswa

4. Menghitung bobot Term Frequency (Tf) Kemunculan kalimat antara Kunci jawaban dengan kelima jawaban siswa.

Tabel 6. Bobot Term frequency

Daftar Kata	Q	D1	D2	D3
lumut	3	3	3	6
hati	1	1	1	2
tanduk	1	1	1	1
batang	2	0	1	2
buluh	2	0	0	0
angkut	2	0	0	0
daun	1	1	1	1
sederhana	1	0	0	0
tunggal	1	0	0	0
lembek	0	0	0	0
struktur	0	3	0	0
kuat	0	0	0	0
bentuk	0	1	2	3
lembar	0	1	1	1

kapsul	0	1	1	0
anjang	0	1	0	0
milik	0	1	1	0
tegak	0	1	1	1
cabang	0	1	1	0
keras	0	0	0	1
sprofit	0	0	0	1
berbentukkapsul	0	0	0	1
memanjang	0	0	0	1
bercabang	0	0	0	1
temu	0	0	0	1
daerah	0	0	0	1
lembab	0	0	0	1

Keterangan

Q : Kunci jawaban guru

D : Dokumen

5. Langkah terakhir menghitung kemiripan menggunakan Cosine Similarity (1) dari bobot hubungan suatu kata Term Frequency (Tf) yang sudah di dapatkan pada Tabel 6

$$D1 = \frac{(3x3) + (1x1) + (1x1) + (2x0) + (2x0) + (2x0) + (1x1) + (1x0) + (1x0) + (0x0) + \dots}{\sqrt{3^2 + 1^2 + 1^2 + 2^2 + 2^2 + 1^2 + 1^2 + 1^2} x}$$

$$\frac{(0x3) + (0x1) + (0x1) + (0x1) + (0x1) + (0x1) + (0x1) + (0x1)}{\sqrt{3^2 + 1^2 + 1^2 + 1^2 + 3^2 + 1^2 + 1^2 + 1^2 + 1^2 + 1^2}} = 0.480$$

$$D2 = \frac{(3x3) + (1x1) + (1x1) + (2x1) + (2x0) + (2x0) + (1x1) + (1x0) + (1x0) + (0x2) + \dots}{\sqrt{3^2 + 1^2 + 1^2 + 2^2 + 2^2 + 1^2 + 1^2 + 1^2} x}$$

$$\frac{(0x1) + (0x1) + (0x1) + (0x1)}{\sqrt{3^2 + 1^2 + 1^2 + 1^2 + 1^2 + 2^2 + 1^2 + 1^2 + 1^2 + 1^2}} = 0.576$$

$$D3 = \frac{(3x6) + (1x2) + (1x1) + (2x2) + (2x0) + (2x0) + (1x1) + (1x0) + (1x0) + (0x3) + \dots}{\sqrt{3^2 + 1^2 + 1^2 + 2^2 + 2^2 + 1^2 + 1^2 + 1^2} x}$$

$$\frac{(0x1) + (0x1) + (0x1)}{\sqrt{6^2 + 2^2 + 1^2 + 2^2 + 1^2 + 3^2 + 1^2 + 1^2 + 1^2 + 1^2 + 1^2 + 1^2 + 1^2 + 1^2}} = 0.632$$

Dari hasil pengujian metode cosine similarity, metode dapat berjalan pada sistem dan dapat memberikan tingkat nilai kemiripan pada sistem agar tingkat kemiripan tersebut dapat di proses menjadi nilai yang di berikan pada siswa.

Nomor Identitas Siswa : 01  
Nama Siswa : Selvira R Yusran

Soal dan Jawaban Siswa

Soal ujian nomor 1  
Jelaskan perbedaan struktur batang pada lumut hati, lumut tanduk, dan lumut daun!  
Jawaban :  
lumut hati berstruktur lembek, lumut tanduk berstruktur kuat dan lumut daun strukturnya sering tumbuh dibagian luar

Bobot Cosine Similarity  
0.480

Nomor Identitas Siswa : 0004597084  
Nama Siswa : Akbar Arif Silawane

Soal dan Jawaban Siswa

Soal ujian nomor 1  
Jelaskan perbedaan struktur batang pada lumut hati, lumut tanduk, dan lumut daun!  
Jawaban :  
lumut hati berbentuk lembaran lumut tanduk berbentuk seperti kapsul memanjang seperti tanduk lumut daun memiliki batang yang tegak dan bercabang

Bobot Cosine Similarity  
0.577

Nomor Identitas Siswa : 0011052721 Nama Siswa : Poetri Fachnun Ayur veda Salatalohy	
<b>Soal dan Jawaban Siswa</b>	
Soal ujian nomor 1 Jelaskan perbedaan realma banyu pada lamur kat, lamur tanduk, dan lamur dano! Jawaban : lamur tanduk : batangnya lebih keras, berbentuk seperti lamur kat tapi ujungnya berstrikapil menatang lamur kat. lamur dengan bentuk benaras yang dimakan diarah benay, lamur dano : lamur anu mampunya bentuk banyu yang tegak dan berabang	Bobot Cosine Similarity 0.632

Gambar 12. Hasil Bobot Metode Cosine Similarity Sistem

#### 4.9. Pengujian Tingkat Akurasi Sistem

Dari hasil pengujian yang di lakukan pada SMA Negeri 11 AMBON, menggunakan mata pelajaran Biologi, dengan total siswa 10 masing-masing siswa mengikuti ujian Biologi, dan setiap mata pelajaran mengandung 1 butir soal. Total item jawaban adalah 10 jawaban siswa di bandingkan dengan kunci jawaban guru, dapat di simpulkan bahwa sistem mempunyai akurasi 80% dalam memberikan nilai dari data yang sudah di dapatkan selama pengujian menggunakan *confision matrix*.

Tabel 7. Confision matrix hasil penelitian

		Nilai sebenarnya	
		True	False
Nilai Prediksi	True	8	1
	False	1	0

Dari tabel 4.1, diperoleh nilai TP = 8, TN = 1, FP = 1, FN = 0.

Langkah berikutnya yaitu dengan menggunakan persamaan (2), dihitung nilai akurasi sebagai berikut:

$$Akurasi = \frac{8 + 0}{8 + 1 + 0 + 1} \times 100\% = \frac{8}{10} = 80\%$$

Kesimpulan yang diperoleh yaitu, dari 10 data uji diperoleh dari jawaban siswa yang berbeda-beda di bandingkan dengan kunci jawaban akurasi kecocokan kelas sebenarnya terhadap jawaban siswa dengan kunci jawaban guru sebesar 80%.

#### 4.10. Pengujian Pemberian Nilai Oleh Sistem

Pengujian pemberian nilai oleh sistem yang di ujikan memakai satu soal dan satu kunci jawaban guru dari total 10 jawaban berbeda-beda dari siswa.

Tabel 8. Pengujian pemberian nilai

Total jawaban siswa	Nilai yang di berikan sistem
1	48.04
2	57.69
3	63.25
4	47.11
5	56.61
6	64.45
7	67.35
8	59.21
9	41.60
10	61.78
<b>Rata-rata</b>	<b>56.71</b>

Dari hasil pengujian pemberian nilai sistem dapat di simpulkan bahwa nilai dari 10 siswa yang di ujitobakan dengan 1 soal dan kunci jawaban Rata-Rata nilai siswa adalah 56.71. Hasil pemberian nilai oleh sistem ini menggunakan nilai bobot Cosine Similarity di kalikan 100 agar bisa mendapatkan nilai sesuai standar penilaian Permendikbud No. 66 tahun 2013 tentang Standar Penilaian Pendidikan (Kurikulum 2013) sesuai dengan sistem seperti pada Gambar 4.12

Nisn	Nama Siswa	Nilai	Aksi
01	Selvira R Yusran	48.04	Lihat Pekerjaan
0004597884	Akbar Arif Silawane	57.69	Lihat Pekerjaan
0011052721	Poetri Fachnun Ayur veda Salatalohy	63.25	Lihat Pekerjaan
0011415519	Muhammad Arya Makasar	47.11	Lihat Pekerjaan
0011612220	Nadiyah Nur Adilah Selayar	56.61	Lihat Pekerjaan
0013454519	Shilya Sarnawia Tusikal	64.45	Lihat Pekerjaan
0013970281	Fadilah Safitri Muliadi Kaimudin	67.35	Lihat Pekerjaan
002022832	Dwi Febriani	59.21	Lihat Pekerjaan
0020290160	Dhea Humairah Faradilla Latar	41.60	Lihat Pekerjaan
0026515528	Isminur Syahirah Marasabessy	61.78	Lihat Pekerjaan

Gambar 13. Hasil pemberian nilai oleh sistem

#### 4.11. Pengujian Fungsional Web

Pengujian fungsional *website* ini dilakukan agar bisa mengetahui apakah sistem sudah bisa berjalan dengan baik pada saat penelitian.

Tabel 9. Fungsional website

No	Menu	Keterangan
1	Halaman Login	Berjalan
2	Halaman Input soal dan jawaban	Berjalan
3	Halaman daftar ujian siswa	Berjalan
4	Halaman input jawaban siswa	Berjalan
5	Halaman Monitoring nilai siswa	Berjalan
6	Halaman formulir pembuatan soal	Berjalan
7	Halaman tingkat kemiripan dan nilai ujian siswa	Berjalan

Pada Tabel 4 menunjukan bahwa sistem penilaian sudah berjalan dengan baik dan tidak ada kendala saat pengujian berlangsung.

### 5. KESIMPULAN DAN SARAN

#### 5.1. Kesimpulan

Setelah dilakukan pengujian sistem Penerapan Metode Cosine Similarity Untuk Pengecekan Kemiripan Jawaban Ujian Siswa pada SMA Negeri 11 Ambon maka penulis mendapatkan kesimpulan sebagai berikut:

1. Pengujian fungsional web dapat berjalan 100% pada dua browser yang di ujitobakan yaitu Google Chrome dan Mozilla Firefox
2. Dalam penelitian ini akurasi sistem dalam memberikan nilai kepada siswa mencapai 80%

berdasarkan data yang di hitung menggunakan Confision matrix

3. Hasil dari pemberian nilai oleh sistem yang di ujicobakan dengan 10 siswa menggunakan 1 soal dan kunci jawaban Rata-Rata nilai dari ke 10 siswa tersebut adalah 56.71.

## 5.2. Saran

Sistem penerapan metode cosine similarity untuk pengecekan jawaban siswa terdapat kekurangan dan kelebihan sehingga di butuhkan saran untuk perbaikan dan penelitian lebih lanjut, berikut ini adalah saran untuk untuk penelitian yang telah di lakukan:

1. Penulis berharap untuk kunci jawaban, di usahakan memberikan jawaban yang sangat berkaitan dengan kata kunci yang di maksud dan tidak terlalu banyak kata-kata yang tidak mengandung arti (bias).
2. Sistem ini dapat di kembangkan lebih lanjut dengan menambahkan metode Jacard, dan algoritma similaritas wu dan palmer agar dapat melakukan proses pengecekan yang lebih baik lagi, agar dapat memberikan nilai yang lebih objektif.

## DAFTAR PUSTAKA

- [1] R. Fitri and A. N. Asyikin, "APLIKASI PENILAIAN UJIAN ESSAY OTOMATIS MENGGUNAKAN," Jurnal POROS TEKNIK, vol. VII, pp. 88-94, 2015.
- [2] K. Rinatha, "Pemodelan Penilaian Essay Otomatis Secara Realtime Menggunakan Kombinasi Text Stemming Dan Cosine Similarity," Konferensi Nasional Sistem & Informatika 2017, p. 322, 2017.
- [3] K. R. Prilianti and H. Wijaya, "Aplikasi Text Mining untuk Automasi Penentuan Tren Topik Skripsi dengan Metode K-Means Clustering," Jurnal Cybermatika, vol. II, pp. 1-2, 2014.
- [4] B. R. Aditya, "Penggunaan Web Crawler Untuk Menghimpun Tweets dengan Metode Pre-Processing Text Mining," Jurnal Infotel Vol. 7 No. 2 November 2015, vol. VII, pp. 95-96, 2015.
- [5] F. Nurhuda, S. W. Sihwi and A. Doewes, "Analisis Sentimen Masyarakat terhadap Calon Presiden Indonesia 2014 berdasarkan Opini dari Twitter Menggunakan Metode Naive Bayes Classifier," JURNAL ITSMART ISSN : 2301-7201, vol. II, pp. 35-42, 2013.
- [6] A. Kurniawan, F. Solihin and F. Hastarita, "PERANCANGAN DAN PEMBUATAN APLIKASI PENCARIAN INFORMASI BEASISWA DENGAN MENGGUNAKAN COSINE SIMILARITY," Jurnal Simantec, vol. IV, pp. 115-124, 2014.
- [7] V. Indriyono, E. Utami and A. Sunyoto, "Pemanfaatan Algoritma Porter Stemmer Untuk Bahasa Indonesia Dalam Proses Klasifikasi Jenis Buku," Jurnal Buana Informatika, vol. VI, pp. 301-309, 2015.
- [8] D. A. R. Ariantini, A. S. M. Lumenta and A. Jacobus, "PENGUKURAN KEMIRIPAN DOKUMEN TEKS BAHASA INDONESIA MENGGUNAKAN METODE COSINE SIMILARITY," E-Journal Teknik Informatika Volume 9, No 1 (2016), ISSN : 2301-8364, vol. IX, pp. 1-8, 2016.
- [9] Irawan, J.D., Andriantantri, E., Trirensila, Y.A., Hasan, M.Z., Yudistira, V.T. and Harhab, K., 2016. Aplikasi Test Online Pada Sd Laboratorium Um Malang. Industri Inovatif Jurnal Teknik Industri, 6(1).R. Intan and A. Defeng, "Hard: Subject-Based Search Engine Menggunakan Tf-Idf Dan Jaccard's Coefficient," Jurnal Teknik Industri Vol. 8, No. 1, JUNI 2006:, pp. 61-72, 2006.
- [10] R. Sovia and J. Febio, "MEMBANGUN APLIKASI E-LIBRARY MENGGUNAKAN HTML, PHP SCRIPT, DAN MYSQL DATABASE," Jurnal PROCESSOR Vol. 6, No.2, Agustus 2011 – STIKOM Dinamika Bangsa - Jambi, vol. VI, pp. 38-54, 2011.
- [11] D. Rahmadiansyah and D. Irwan, "mplementasi Metode Model View Controller Menggunakan Framework Code Igniter dalam Pengembangan Aplikasi Manajemen Depo Petikemas pada Unit Usaha Belawan Logistics Center," Seminar Nasional Teknologi Informasi dan Komunikasi (SNASTIKOM 2012) ISBN 978-602-19837-0-6, pp. 1-11, 2012.
- [12] A. Indriani, "Klasifikasi Data Forum dengan menggunakan Metode Naïve Bayes Classifier," Seminar Nasional Aplikasi Teknologi Informasi (SNATI) ISSN: 1907 - 5022, pp. G5-G10, 2014.