



Penerapan Algoritma *Stemming Nazief-Adriani* dengan Metode *Cosine Similarity* Dalam Aplikasi Ujian Esai

Mohamad Arif Suryawan^{1*}, LM. Fajar Israwan², Ferdianto Arland³

^{1,2,3}Program Studi Teknik Informatika, Fakultas Teknik, Universitas Dayanu Ikhsanuddin
arwan97@unidayan.ac.id

Abstract

Essay exam assessment using an exam application takes a long time to check. Answer checking is done by reading carefully. The answers given are some that match the answer key, and some are very different. Therefore, an exam application is needed that can directly assess the essay question answers and then give a score according to the answers written. This study aims to apply the Nazief-Adriani stemming algorithm with the Cosine Similarity method to the high school essay exam application. The research method used in this study consists of two stages, first the Nazief-Adriani algorithm carries out several processes, namely normalizing lowercase letters, punctuation, separating into individual words, removing common words, stemming by removing affixes from root words, mapping synonyms and finally calculating the frequency of word occurrence. Furthermore, the second stage, the Cosine Similarity method is used to compare the level of similarity with the answer key. This study produces an exam application that applies the Nazief-Adriani stemming algorithm in checking essay answers. This exam application is Android-based so that it makes it easier to answer exam questions, the answers are written in the box provided. The essay exam questions that are given are first equipped with the answer key stored in the application. From testing three essay answers, the calculation results using the Cosine Similarity method were obtained, namely: the first answer is 94.3, the second answer is 94.3, and the third answer is 74.5. The first and second answers produce a value of 94.3, indicating a high level of similarity to the answer key. Thus, the application created is expected to make it easier to check answers to essay questions quickly and accurately.

Keywords: application, cosine similarity, essay, nazief-adriani, stemming, exam

Abstrak

Penilaian ujian esai dengan menggunakan aplikasi ujian membutuhkan waktu lama dalam pemeriksaan. Pemeriksaan jawaban dilakukan dengan cara membaca dengan teliti. Jawaban yang diberikan ada yang sesuai dengan kunci jawaban, ada juga yang berbeda jauh. Oleh karena itu dibutuhkan sebuah aplikasi ujian yang dapat menilai secara langsung dari jawaban soal esai kemudian memberikan nilai sesuai dengan jawaban yang ditulis. Penelitian ini bertujuan untuk menerapkan algoritma *stemming Nazief-Adriani* dengan metode *Cosine Similarity* pada aplikasi ujian esai tingkat SMA. Metode penelitian yang digunakan dalam penelitian ini terdiri dari dua tahap, pertama algoritma *Nazief-Adriani* melakukan beberapa proses yaitu normalisasi huruf kecil, tanda baca, memisahkan dalam kata-kata individual, menghapus kata umum, *stemming* dengan cara menghilangkan imbuhan dari kata dasar, memetakan kata sinonim dan terakhir menghitung frekuensi kemunculan kata. Selanjutnya tahap kedua, metode *Cosine Similarity* digunakan untuk membandingkan tingkat kemiripan dengan kunci jawaban. Penelitian ini menghasilkan aplikasi ujian yang menerapkan algoritma *stemming Nazief-Adriani* dalam memeriksa jawaban esai. Aplikasi ujian ini berbasis android sehingga memudahkan dalam menjawab soal ujian, jawaban dituliskan dalam kotak yang disiapkan. Soal ujian esai yang diberikan terlebih dahulu dilengkapi dengan kunci jawaban yang disimpan dalam aplikasi tersebut. Dari pengujian tiga jawaban esai didapatkan hasil perhitungan dengan metode *Cosine Similarity* yaitu: jawaban pertama adalah 94,3 jawaban kedua adalah 94,3 dan jawaban ketiga adalah 74,5. Jawaban pertama dan kedua menghasilkan nilai 94,3 menunjukkan tingkat kemiripan yang tinggi terhadap kunci jawaban. Dengan demikian aplikasi yang dibuat diharapkan dapat memudahkan dalam memeriksa jawaban dari soal esai secara cepat dan tepat.

Kata kunci: aplikasi, cosine similarity, esai, nazief-adriani, stemming, ujian

1. Pendahuluan

Pendidikan adalah serangkaian proses yang berisikan teknik dan metode belajar mengajar sesuai dengan standar yang telah ditetapkan yang bertujuan untuk menyebarkan pengetahuan. Pendidikan memiliki peran

yang penting dalam membentuk karakter pelajar. Proses pendidikan di sekolah dapat diukur, salah satunya dengan evaluasi hasil belajar. Salah satu cara melakukan evaluasi yaitu dengan memberi ujian dalam bentuk pilihan ganda dan esai.

Ujian esai adalah salah satu bentuk dari evaluasi untuk mengetahui kemampuan siswa terhadap materi yang telah diberikan. Jika dibandingkan dengan ujian dalam bentuk pilihan ganda, ujian esai membutuhkan waktu lama dalam menjawab soal. Evaluasi dengan memberikan soal esai kepada peserta didik dilakukan dengan memberikan soal pada lembaran kertas kemudian dijawab pada lembar jawaban. Selanjutnya pengajar harus memeriksa jawaban tersebut secara manual. Penilaian hasil ujian yang dilakukan secara manual memiliki kelemahan, yaitu pengajar membutuhkan waktu lama untuk memeriksa jawaban. Semakin banyak jumlah soal dan siswa yang mengikuti ujian, semakin banyak pula lembar jawaban yang harus diperiksa. Hal ini tentu saja menyebabkan kualitas penilaian menurun dan kualitas penilaian tidak lagi objektif.

Beberapa penelitian yang pernah dilakukan sebelumnya dan menjadi referensi penelitian ini, antara lain dengan judul Algoritma Bastal: Adaptasi Algoritma Nazief & Adriani untuk *Stemming* Teks Bahasa Bali. Penelitian ini bertujuan untuk menggunakan metode *stemming* untuk menentukan kata dasar pada teks bahasa Bali yang diberi nama algoritma Bastal. Hasil validasi menunjukkan bahwa aplikasi mampu bekerja dengan baik. Aplikasi dinilai sangat akurat dengan persentase keseluruhan penilaian yaitu 96,15%. Aplikasi juga mendapatkan respon positif dari pengguna [1].

Penelitian berikutnya dengan judul Sistem Penilaian Esai Otomatis Menggunakan Algoritma *Stemming* Nazief dan Adriani. Penelitian ini bertujuan untuk membuat *E-Learning* dengan fitur sistem penilaian otomatis menggunakan *stemming* Nazief & Adriani. Hasil dari penelitian proses *Information Retrieval* (IR) untuk menghitung bobot menggunakan *frequency accuracy* yang lebih rendah, dengan akurasi tinggi 90,66% jika dibandingkan dengan menggunakan perhitungan *Inverse Document Frequency* (IDF) [2].

Selanjutnya penelitian dengan judul Algoritma *Stemming* Nazief & Adriani dengan Metode *Cosine Similarity* untuk Chatbot Telegram Terintegrasi dengan E-layanan. Penelitian ini bertujuan untuk menghitung tingkat akurasi kesesuaian respons chatbot terhadap pertanyaan yang masuk dengan algoritma *stemming* Nazief & Adriani dan metode kemiripan menggunakan *Cosine Similarity*. Dengan mengimplementasikan algoritma dan metode tersebut dalam chatbot dengan bantuan layanan Messenger Telegram menghasilkan jawaban yang relatif sesuai dengan yang diharapkan pengguna. Cara ini merupakan cara yang efektif untuk menjawab pertanyaan secara otomatis. Dan jika pertanyaan yang diajukan tidak menemukan jawaban atau jawaban tidak sesuai maka pertanyaan akan di sampaikan [3].

Penelitian lainnya dengan judul Pembuatan Aplikasi Ujian Siswa di SMA Berbasis Web. Penelitian ini bertujuan untuk memberikan kemudahan untuk

pelaksanaan ujian dan pemeriksaan lembar jawaban siswa. Dilihat dari hasil jawaban siswa untuk soal pilihan ganda otomatis terperiksa di aplikasi sedangkan jawaban esai akan diperiksa secara langsung oleh guru yang bersangkutan melalui aplikasi. Hasil ujian dapat langsung di cetak dari aplikasi tersebut [4].

Penelitian dengan judul Pemanfaatan *Vector Space Model* Algoritma Nazief Adriani Pembobotan TFIDF Pada *Prototype* Klasifikasi Teks Bahasa Indonesia. Penelitian ini bertujuan untuk menguji aplikasi mesin klasifikasi teks bahasa Indonesia menggunakan algoritma *Stemming* Nazief Adriani, K-Algoritma *Nearest Neighbor* dan Model Ruang Vektor berdasarkan pembobotan frekuensi TFIDF jumlah kata dan fungsi *Simpson*. Dengan menggunakan *document news* sebagai pembelajaran dokumen sebanyak 15 (lima belas) dokumen dengan 3 (tiga) kategori, menghasilkan nilai rata-rata *Precision* dan *Recall* sebesar 81,33% [5].

Selanjutnya penelitian lainnya dengan judul Deteksi Lembar Jawaban Komputer Menggunakan OMR (*Optical Mark Recognition*) di MTs Nurul Iman. Penelitian ini bertujuan untuk mengembangkan suatu metode yang secara otomatis mendeteksi pilihan jawaban pada lembar jawaban komputer, sehingga diperoleh hasil yang lebih akurat dan cepat. Dilihat dari hasil tingkat akurasi deteksi lembar jawaban komputer dengan menggunakan OMR adalah 97% [6].

Penelitian dengan judul Pengaruh *Stemming* Nazief & Adriani terhadap Performa Algoritma *RabinKarp* dalam Mendeteksi Kemiripan Teks. Penelitian ini bertujuan untuk menganalisis terhadap pengaruh metode *stemming* Nazief Adriani terhadap kinerja algoritma *Rabin-Karp* dalam mendeteksi tingkat kemiripan teks. Dilihat dari Hasil validasi menunjukkan bahwa data yang mengalami perubahan susunan kata secara acak yang mengimplementasikan metode *stemming* mengalami rata-rata penurunan sebesar 0,76% dibandingkan tanpa implementasi metode *stemming* [7].

Penelitian lainnya dengan judul *Automated Essay Scoring* menggunakan *Cosine Similarity* pada Penilaian Esai Multi Soal. Penelitian ini bertujuan untuk menghasilkan Implementasi *Automated Essay Scoring* (AES) agar dapat mempermudah dalam melakukan koreksi dan pemberian skor dengan memanfaatkan aplikasi yang dijalankan pada komputer. Dengan menggunakan prinsip similaritas maka hanya karakter teks yang relevan dari dokumen ahli yang akan diberi bobot, untuk mendapatkan nilai yang tepat kemiripannya, jumlah kata yang terkandung pada dokumen jawaban harus sesuai, tidak berjumlah lebih sedikit dan juga tidak berjumlah lebih banyak [8].

Penelitian dengan judul Implementasi Algoritma *Text Mining* dan *Cosine Similarity* untuk Desain Sistem Aspirasi Publik Berbasis Mobile. Penelitian ini bertujuan untuk membangun sebuah aplikasi yang

dapat menampung aspirasi masyarakat dan kemudian dapat mengklasifikasikannya dan meneruskannya ke komisi berbasis mobile yang sesuai. Dilihat dari Hasil validasi menunjukkan bahwa Aplikasi menggunakan kombinasi metode *text mining* dan *Cosine similarity* mengukur kesamaan fungsional masing-masing komisi dengan aspirasi yang disuarakan oleh masyarakat sehingga aspirasi tepat sasaran [9].

Penelitian lainnya dengan judul Representasi Teknik *Information Retrieval* Pada Perhitungan *Rabin-Karp* Menggunakan *Stemming Nazief-Adriani*. Penelitian ini bertujuan untuk menerapkan sistem *information retrieval* dalam *text mining*, yaitu yang berhubungan dengan tingkat kemiripan informasi. Dilihat dari Hasil validasi menunjukkan bahwa model mampu untuk menghitung tingkat kemiripan teks [10].

Metode *cosine similarity* digunakan untuk memeriksa kemiripan teks kata kunci dalam kunci jawaban. Hal ini sangat mempengaruhi hasil perhitungan untuk menentukan penilaian [11]. Sebelum proses pemeriksaan kemiripan teks, dilakukan seleksi dalam sistem dengan proses *stemming*. Proses tersebut mengubah kata-kata yang terdapat dalam kalimat menjadi kata dasar dengan menerapkan aturan tertentu [12]. *Stemming* dalam bahasa Indonesia merupakan tahapan yang menghilangkan sufiks, prefiks dan awalan [13]. Tahapan dalam proses *stemming* tersebut akan memudahkan dalam proses pencocokan dengan kunci jawaban.

Penerapan *stemming* pada kata-kata dari sebuah kalimat dapat membaca afiks menjadi kata dasar. Kata dasar tersebut diawali dengan menghilangkan afiks dari akhiran dan awalan menjadi kata dasar. Algoritma *Nazief-Adriani* dengan model *finite state automata* dapat diimplementasikan dengan sistem berbasis web yang dapat menghasilkan model penyelesaian kata afiks menjadi kata dasar bahasa Indonesia [14].

Algoritma *Nazief-Adriani* berdasarkan hasil pengujian memperlihatkan keberhasilan lebih baik, nilai rata-rata akurasi tertinggi yakni sebesar 97.73% dengan rata-rata waktu proses *stemming* selama 20.17 detik. Sedangkan menggunakan algoritma Arifin Setiono lebih rendah performansinya dengan nilai rata-rata akurasi sebesar 94.37% dengan rata-rata waktu proses *stemming* selama 23.32 detik [15].

Pengembangan penelitian selanjutnya dengan judul Penerapan Algoritma *Stemming Nazief-Adriani* dengan Metode *Cosine Similarity* Dalam Aplikasi Ujian Esai. Penelitian ini bertujuan untuk menerapkan algoritma *stemming Nazief-Adriani* dengan metode *Cosine Similarity* pada aplikasi ujian esai tingkat SMA. Aplikasi ini diharapkan dapat membantu proses penilaian jawaban peserta ujian esai secara otomatis dan akurat.

2. Metode Penelitian

2.1 Algoritma *Nazief-Adriani*

Algoritma *Nazief-Adriani* memiliki beberapa tahapan yaitu:

Normalisasi Huruf Kecil: Semua huruf diubah menjadi huruf kecil untuk menjaga konsistensi. Contoh: "Belajar" dan "belajar" menjadi "belajar".

Penghapusan Tanda Baca: Semua tanda baca dihapus untuk memastikan hanya kata yang tersisa. Contoh: "Halo, dunia!" menjadi "Halo dunia".

Tokenisasi: Teks dipecah menjadi kata-kata individual. Contoh: "Saya belajar pemrograman" menjadi ["Saya", "belajar", "pemrograman"].

Penghapusan Kata Umum: Kata-kata yang sering muncul tapi tidak penting dihapus. Contoh: "Saya belajar di rumah" menjadi "belajar rumah".

Stemming: Imbuhan dihilangkan untuk mendapatkan akar kata. Contoh: "belajar", "belajaran", dan "mempelajari" menjadi "ajar".

Pemetaan Sinonim: Kata-kata dengan makna yang sama disatukan. Contoh: "indah" dan "cantik" bisa dipetakan menjadi satu kata.

Penghitungan Frekuensi Kata: Frekuensi kemunculan setiap kata dihitung. Contoh: kata "pemrograman" muncul 10 kali berarti frekuensinya 10.

2.2 Metode Cosine Similarity

Metode *cosine similarity* merupakan metode yang digunakan untuk menghitung tingkat kesamaan (*similarity*) antar dua buah objek yang dinyatakan dalam dua buah vektor dengan menggunakan kata kunci (*keyword*) dari sebuah dokumen sebagai ukuran. Tingkat *similarity* pada fungsi *cosine similarity* berkisar pada interval 0 (nol) sampai dengan 1 (satu). Nilai 0 (nol) melambangkan kedua objek sama sekali berbeda, sedangkan nilai 1 (satu) melambangkan kedua objek persis atau sama

Untuk mendapatkan hasil dari fungsi tersebut dapat dilakukan normalisasi. Berikut Persamaan 1 yang digunakan untuk menghitung tingkat kesamaan pada *cosine similarity*.

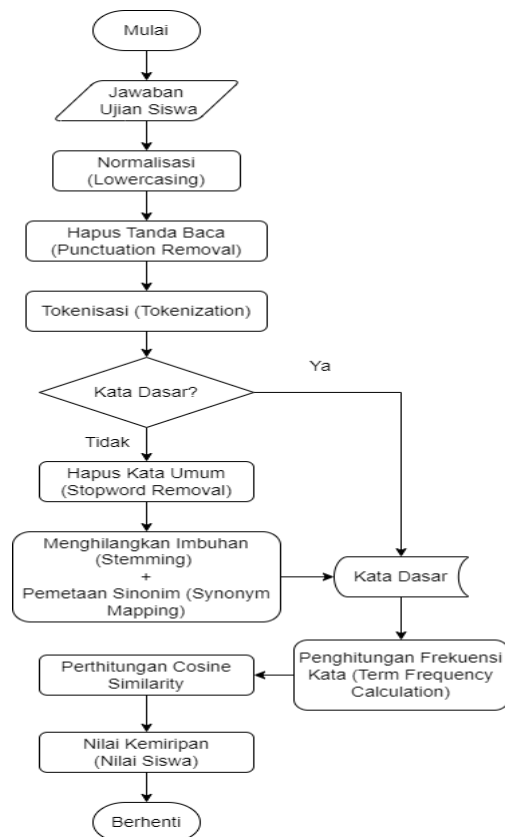
$$\text{Cosine Similarity} = \frac{X * Y}{\sqrt{|X|} * \sqrt{|Y|}} * 100 \quad (1)$$

X.Y adalah X dan Y dihitung dengan $\sum_k^n = X_k Y_k$, |X| adalah panjang vektor x, dihitung dengan $\sum_k^n = X_k^2$, |Y| adalah panjang vektor y, dihitung dengan $\sum_k^n = Y_k^2$.

2.3 Penilaian Esai

Penilaian soal esai menggambarkan alur proses pemeriksaan jawaban peserta ujian terhadap kunci jawaban yang telah disiapkan. Proses dimulai memeriksa jawaban peserta ujian. Jawaban tersebut dinormalisasi dengan Algoritma *Nazief-Adriani* dengan melakukan proses perubahan yaitu: substitusi, penyisipan, dan penghapusan karakter. Setelah kata terbentuk, selanjutnya perhitungan cosine similarity.

Proses ini akan membandingkan berapa kata yang sama dan kemudian dihitung nilai kemiripannya. *Flowchart* pada Gambar 1, akan memberikan tahapan yang berjalan dalam sistem penilaian.



Gambar 1. *Flowchart* Penilaian Esai

3. Hasil dan Pembahasan

Hasil dan pembahasan dari aplikasi ujian yang menggunakan algoritma *Nazief-Adriani* dan metode *cosine similarity* dapat dilihat dari pengujian berikut. Pengujian ini dilakukan secara manual terlebih dahulu untuk kemudian dibandingkan dengan hasil penilaian dari sistem. Dengan membandingkan kedua hasil ini, kita dapat mengevaluasi keakuratan dan konsistensi sistem penilaian otomatis yang telah dikembangkan. Gambar 2, 3 dan 4 adalah Input Jawaban Soal Esai.



Gambar 2. Jawaban Esai Pertama



Gambar 3. Jawaban Esai Kedua



Gambar 4. Jawaban Esai Ketiga

Tabel 1 dan 2 adalah proses pemeriksaan jawaban menggunakan algoritma *Nazief-Adriani* dan *cosine similarity*:

Tabel 1. Kunci Jawaban

Pertanyaan	Kunci Jawaban
Apa yang dimaksud dengan poster?	Poster adalah media publikasi yang menggabungkan teks dan gambar untuk menyampaikan informasi atau pesan kepada publik

Tabel 2. Jawaban Soal Esai

Esai	Jawaban
Pertama	Poster adalah media publikasi yang menggabungkan teks dan gambar untuk menyampaikan informasi atau pesan kepada publik
Kedua	Media publikasi poster mengandung teks dan gambar untuk menyampaikan informasi, pesan, atau ajakan kepada masyarakat luas
Ketiga	Poster berisi teks dan gambar yang digunakan untuk menyampaikan informasi atau pesan kepada orang banyak

Lowercasing adalah proses standarisasi teks dengan mengubah semua huruf menjadi huruf kecil. Hal ini dilakukan untuk memudahkan analisis teks, karena dengan huruf kecil, kata-kata yang sebenarnya sama namun ditulis dengan huruf besar dan kecil dianggap sama.

Tabel 3. *Lowercasing* Jawaban

Esai	Jawaban	11	gambar	menyampaika	pesan	atau
Kunci	poster adalah media publikasi yang menggabungkan teks dan gambar untuk menyampaikan informasi atau pesan kepada publik	12	yang	informasi	atau	pesan
Pertama	poster adalah media publikasi yang menggabungkan teks dan gambar untuk menyampaikan informasi atau pesan kepada publik	13	digunakan	atau	ajakan	kepada
Kedua	media publikasi poster mengandung teks dan gambar untuk menyampaikan informasi, pesan, atau ajakan kepada masyarakat luas	14	untuk	pesan	kepada	orang
Ketiga	poster berisi teks dan gambar yang digunakan untuk menyampaikan informasi atau pesan kepada orang banyak	15	menyampai	kepada	masyarakat	banyak
		16	informasi	publik	luas	
		17	pesan			
		18	atau			
		19	ajakan			
		20	kepada			
		21	khalayak			
		22	umum			

Punctuation Stripping adalah proses menghilangkan tanda baca (*punctuation*) dari teks. Tanda baca seperti titik, koma, tanda tanya, tanda seru, dan tanda baca lainnya dihapus dari teks. Tujuan dari *Punctuation Stripping* adalah untuk memproses teks sehingga hanya sisa kata-kata atau token yang menjadi fokus analisis, tanpa gangguan dari tanda baca yang tidak relevan seperti pada Tabel 4.

Tabel 4. *Punctuation Stripping* Jawaban

Aktor	Jawaban
Kunci	poster adalah media publikasi yang menggabungkan teks dan gambar untuk menyampaikan informasi atau pesan kepada publik
Pertama	poster adalah media publikasi yang menggabungkan teks dan gambar untuk menyampaikan informasi atau pesan kepada publik
Kedua	media publikasi poster mengandung teks dan gambar untuk menyampaikan informasi pesan atau ajakan kepada masyarakat luas
Ketiga	poster berisi teks dan gambar yang digunakan untuk menyampaikan informasi atau pesan kepada orang banyak

Tokenisasi adalah proses memecah teks atau kalimat menjadi bagian-bagian yang lebih kecil, yang disebut sebagai token. Token dapat berupa kata-kata, frasa, simbol, atau bahkan karakter, tergantung pada jenis tokenisasi yang dilakukan. Proses ini membantu dalam memahami struktur teks dan memungkinkan analisis lebih lanjut, seperti pemrosesan bahasa alami, analisis teks, dan pengindeksan informasi seperti pada Tabel 5.

Tabel 5. Tokenisasi Kunci Jawaban

No	Kunci	Pertama	Kedua	Ketiga
1	poster	poster	media	poster
2	adalah	adalah	publikasi	berisi
3	sebuah	media	poster	teks
4	media	publikasi	mengandung	dan
5	publikasi	yang	teks	gambar
6	yang	menggabungkan	dan	yang
7	terdiri	teks	gambar	digunakan
8	dari	dan	untuk	untuk
9	teks	gambar	menyampaikan	menyampaikan
10	dan	untuk	informasi	informasi

Filtering Stopwords adalah proses menghapus kata-kata yang umum dan tidak memberikan makna tambahan dalam teks. Kata-kata tersebut disebut *stopwords*. Contoh *stopwords* adalah "dan", "atau", "yang", dan lain-lain seperti pada Tabel 6.

Tabel 6. *Filtering Stopwords* Kunci Jawaban

No	Kunci	Pertama	Kedua	Ketiga
1	poster	poster	media	poster
2	adalah	adalah	publikasi	berisi
3	sebuah	media	poster	teks
4	media	publikasi	mengandung	dan
5	publikasi	yang	teks	gambar
6	yang	menggabungkan	dan	yang
7	terdiri	teks	gambar	digunakan
8	dari	dan	untuk	untuk
9	teks	gambar	menyampaikan	menyampaikan
10	dan	untuk	informasi	informasi
11	gambar	menyampaikan	pesan	atau
12	yang	informasi	atau	pesan
13	digunakan	atau	ajakan	kepada
14	untuk	pesan	kepada	orang
15	menyampaikan	kepada	masyarakat	banyak
16	informasi	publik	luas	
17	pesan			
18	atau			
19	ajakan			
20	kepada			
21	khalayak			
22	umum			

Kata-kata yang berwarna abu-abu pada tabel diatas berisi daftar *stopword* yang telah dihapus dari data *array* tokenisasi masing-masing jawaban. Proses ini dilakukan untuk memastikan bahwa analisis teks dapat dilakukan dengan lebih akurat, karena *stopword* adalah kata-kata umum yang biasanya tidak memberikan nilai informatif signifikan dalam analisis teks.

Stemming adalah proses menghilangkan awalan (*prefix*) dan akhiran (*suffix*) pada kata dalam teks untuk menghasilkan bentuk dasar atau kata dasar (*root word*). Proses ini bertujuan untuk menyederhanakan kata-kata menjadi bentuk dasarnya, sehingga memudahkan analisis dan pemrosesan teks. Dalam pengolahan bahasa alami (NLP), *stemming* sangat penting untuk

mengurangi variasi kata yang muncul akibat infleksi atau perubahan bentuk kata seperti pada Tabel 7.

Tabel 7. Stemming Kunci Jawaban

No	Kunci	Pertama	Kedua	Ketiga
1	poster	poster	media	poster
2	media	media	publikasi	isi
3	publikasi	publikasi	poster	teks
4	teks	gabung	kandung	gambar
5	gambar	teks	teks	informasi
6	informasi	gambar	gambar	pesan
7	pesan	informasi	informasi	orang
8	ajak	pesan	pesan	
9	khalayak	khalayak	ajak masyarakat luas	

Kata-kata pada Tabel 7 yang berwarna abu-abu telah dihilangkan imbuhan untuk menjadikannya sebagai kata dasar yang murni, sehingga memudahkan dalam analisis linguistik dan pemahaman makna kata tersebut.

Indeksasi dan *Cosine Similarity*: Indeksasi adalah proses membuat indeks atau daftar kata-kata yang ada dalam sebuah teks beserta dengan lokasi atau informasi terkait kata-kata tersebut. Indeks ini digunakan untuk mempercepat pencarian kata-kata atau frasa dalam teks. Dalam konteks pemrosesan teks, indeksasi sering kali mengacu pada proses membuat indeks dari koleksi dokumen teks, sehingga memungkinkan pencarian informasi yang efisien dalam dokumen-dokumen tersebut.

Tabel 8. Proses Indeksasi Jawaban Pertama, Kedua dan Ketiga

No	Jawaban	X	Y	X ²	Y ²	X.Y
1	poster	111	111	111	111	111
2	media	111	110	111	110	110
3	publikasi	111	110	111	110	110
4	teks	111	111	111	111	111
5	gambar	111	111	111	111	111
6	informasi	111	111	111	111	111
7	pesan	111	111	111	111	111
8	ajakan	111	010	111	010	010
9	khalayak	111	100	111	100	100
Total				999	885	885

Selanjutnya perhitungan nilai jawaban pertama menggunakan metode *cosine similarity*:

$$\begin{aligned}
 X * Y &= (1 \times 1) + (1 \times 1) + (1 \times 1) + (1 \times 1) + (1 \times 1) + \\
 &\quad (1 \times 1) + (1 \times 1) + (1 \times 0) + (1 \times 1) \\
 &= (1+1+1+1+1+1+1+0+1) = 8 \\
 |X| &= (1^2+1^2+1^2+1^2+1^2+1^2+1^2+1^2) \\
 &= (1+1+1+1+1+1+1+1) \\
 &= 9 \\
 |Y| &= (1^2+1^2+1^2+1^2+1^2+1^2+0^2+1^2) \\
 &= (1+1+1+1+1+1+1+0+1) \\
 &= 8 \\
 \text{Nilai} &= \frac{8}{\sqrt{9} * \sqrt{8}} * 100 = \frac{8}{3 * 2,828} * 100 = \frac{8}{8,485} \\
 &= 0,943 * 100 = 94,3
 \end{aligned}$$

Selanjutnya, perhitungan nilai jawaban kedua:

$$\begin{aligned}
 X * Y &= (1 \times 1) + (1 \times 1) + (1 \times 1) + (1 \times 1) + (1 \times 1) + \\
 &\quad (1 \times 1) + (1 \times 1) + (1 \times 0) + (1 \times 1) \\
 &= (1+1+1+1+1+1+1+0+1) = 8
 \end{aligned}$$

$$= (1^2+1^2+1^2+1^2+1^2+1^2+1^2+1^2)$$

$$= (1+1+1+1+1+1+1+1) = 9$$

$$|Y| = (1^2+1^2+1^2+1^2+1^2+1^2+0^2+1^2)$$

$$= (1+1+1+1+1+1+1+0+1) = 8$$

$$\text{Nilai} = \frac{8}{\sqrt{9} * \sqrt{8}} * 100 = \frac{8}{3 * 2,828} * 100 = \frac{8}{8,485}$$

$$= 0,943 * 100 = 94,3$$

Terakhir, perhitungan nilai ketiga jawaban ketiga:

$$\begin{aligned}
 X * Y &= (1 \times 1) + (1 \times 0) + (1 \times 0) + (1 \times 1) + (1 \times 1) + \\
 &\quad (1 \times 1) + (1 \times 1) + (1 \times 0) + (1 \times 0) \\
 &= (1+0+0+1+1+1+1+0+0) = 5
 \end{aligned}$$

$$|X| = (1^2+1^2+1^2+1^2+1^2+1^2+1^2+1^2)$$

$$= (1+1+1+1+1+1+1+1) = 9$$

$$|Y| = (1^2+0^2+0^2+1^2+1^2+1^2+0^2+0^2)$$

$$= (1+0+0+1+1+1+1+0+0) = 5$$

$$\text{Nilai} = \frac{5}{\sqrt{9} * \sqrt{5}} * 100 = \frac{5}{3 * 2,236} * 100 = \frac{5}{6,708}$$

$$= 0,745 * 100 = 74,5$$

Perhitungan yang telah dilakukan dari ketiga jawaban esai menggunakan metode *cosine similarity*, didapat nilai untuk jawaban pertama adalah 94,3 jawaban kedua adalah 94,3 dan jawaban ketiga adalah 74,5. Dengan demikian hasil tersebut sesuai dengan nilai yang didapat dalam aplikasi ujian esai seperti terlihat pada Gambar 5, 6 dan 7.

Jawaban :

Poster adalah media publikasi yang menggabungkan teks dan gambar untuk menyampaikan informasi atau pesan kepada publik

Nilai : 94.3

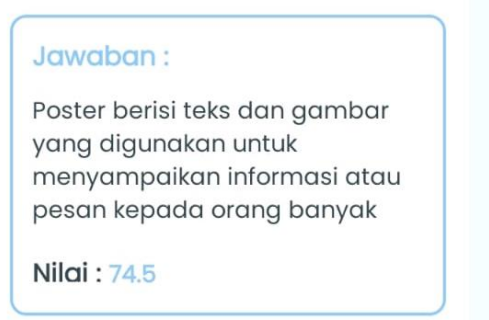
Gambar 5. Nilai Jawaban Esai Pertama

Jawaban :

Media publikasi poster mengandung teks dan gambar untuk menyampaikan informasi, pesan, atau ajakan kepada masyarakat luas

Nilai : 94.3

Gambar 6. Nilai Jawaban Esai Kedua



Gambar 7. Nilai Jawaban Esai Ketiga

4. Kesimpulan

Penelitian ini menghasilkan aplikasi ujian esai tingkat SMA dengan menerapkan algoritma *stemming Nazief-Adriani* dalam memeriksa jawaban esai. Pemeriksaan jawaban dilakukan beberapa tahapan yaitu normalisasi huruf kecil, penghapusan tanda baca, tokenisasi dengan memisahkan menjadi kata-kata individual, penghapusan kata umum, *stemming* dengan menghapus imbuhan kata, pemetaan sinonim dan terakhir menghitung frekuensi kata. Selanjutnya dengan menggunakan metode *Cosine Similarity* menghitung tingkat kesamaan jawaban esai dengan kunci jawaban. Dari pengujian tiga jawaban esai didapatkan hasil perhitungan yaitu: jawaban pertama adalah 94,3 jawaban kedua adalah 94,3 dan jawaban ketiga adalah 74,5. Jawaban pertama dan kedua menghasilkan nilai 94,3 menunjukkan tingkat kemiripan yang tinggi terhadap kunci jawaban. Dengan demikian aplikasi yang dibuat diharapkan dapat digunakan sebagai aplikasi ujian yang memudahkan dalam pelaksanaannya dan pemeriksaan jawaban dari soal esai secara cepat dan tepat.

Daftar Rujukan

- [1] I. P. M. Wirayasa, I. M. A. Wirawan, dan I. M. A. Pradnyana, "Algoritma Bastal: Adaptasi Algoritma Nazief & Adriani Untuk Stemming Teks Bahasa Bali," *j. nas. pendidik. teknik. inform.*, vol. 8, no. 1, hlm. 60, Jun 2019, doi: 10.23887/janapati.v8i1.13500.
- [2] R. Ahmad dan R. R. O. Sasue, "Sistem Penilaian Esai Otomatis Menggunakan Algoritma Stemming Nazief Dan Adriani," *JTTL*, vol. 1, no. 2, hlm. 101–108, 2020.
- [3] A. C. Herlingga, I. P. E. Prisma, D. R. Prehanto, dan D. A. Dermawan, "Algoritma Stemming Nazief & Adriani dengan Metode Cosine Similarity untuk Chatbot Telegram Terintegrasi dengan E-layanan," *JINACS*, vol. 2, no. 01, hlm. 19–26, Okt 2020, doi: 10.26740/jinacs.v2n01.p19-26.
- [4] I. R. Bakti, "Pembuatan Aplikasi Ujian Siswa di SMA Berbasis Web," *SYNTAX*, vol. 2, no. 1, hlm. 82–85, Jul 2021, doi: 10.46576/syntax.v2i1.1264.
- [5] R. Fitrianyah, "Pemanfaatan Vector Space Model Algoritma Nazief Andriani Pembobotan Tfidf Pada Prototipe Klasifikasi Teks Bahasa Indonesia," *JuTech*, vol. 2(1), hlm. 37–47, 2021.
- [6] M. N. Hermawan, "Deteksi Lembar Jawaban Komputer Menggunakan OMR (Optical Mark Recognition) Di MTS Nurul Iman," *JATISI*, vol. 8, no. 3, hlm. 1361–1372, Sep 2021, doi: 10.35957/jatisi.v8i3.1078.
- [7] M. A. Yulianto, "Pengaruh Stemming Nazief & Adriani terhadap Performa Algoritma Rabin- Karp dalam Mendeteksi Kemiripan Teks," vol. 6, no. 4, 2021.
- [8] A. R. Lahitani, "Automated Essay Scoring menggunakan Cosine Similarity pada Penilaian Esai Multi Soal," *J. Kaji. Ilm.*, vol. 22, no. 2, hlm. 107–118, Mei 2022, doi: 10.31599/jki.v22i2.1121.
- [9] R. Rismayani, H. Sy, T. Darwansyah, dan I. Mansyur, "Implementasi Algoritma Text Mining dan Cosine Similarity untuk Desain Sistem Aspirasi Publik Berbasis Mobile," *Komputika*, vol. 11, no. 2, hlm. 169–176, Agu 2022, doi: 10.34010/komputika.v11i2.6501.
- [10] A. C. Sitepu, J. Panjaitan, dan M. Sigi, "Representasi Teknik Information Retrieval Pada Perhitungan Rabin-Karp Menggunakan Stemming Nazief-Adriani," vol. 2, no. 1, 2022.
- [11] M. S. Simanjuntak, J. Panjaitan, dan S. A. Syahputra, "Using Preprocessing Text Mining With Nazief-Adriani Algorithms Similarity Of Essay Final Exam Semester," *Jurnal Mantik*, vol. 4, no. 1, hlm. 1714–1720, 2020.
- [12] A. Prasidhatama dan K. M. Suryaningrum, "Perbandingan Algoritma Nazief & Adriani Dengan Algoritma Idris Untuk Pencarian Kata Dasar," *J. Teknol. dan Manaj. Inform.*, vol. 4, no. 1, hlm. 1–4, 2018, doi: 10.26905/jtmi.v4i1.1773.
- [13] H. T. Nugroho, "Pengaruh Algoritma Stemming Nazief-Adriani Terhadap Kinerja Algoritma Winnowing Untuk Mendeteksi Plagiarisme Bahasa Indonesia," *J. Ultim. Comput.*, vol. 9, no. 9, hlm. 36–40, 2017, doi: 10.31937/sk.v9i1.572.
- [14] L. A. Fitriana, A. Mustopa, M. R. Firdaus, dan R. Dahlia, "Application of the Finite State Automata (FSA) Method in Indonesian Stemming using the Nazief & Adriani Algorithm," *SISTEMASI*, vol. 13, no. 3, hlm. 1125, Mei 2024, doi: 10.32520/stmsi.v13i3.4038.
- [15] A. Sinaga dan S. P. Nainggolan, "Analisis Perbandingan Akurasi Dan Waktu Proses Algoritma Stemming Arifin-Setiono Dan Nazief-Adriani Pada Dokumen Teks Bahasa Indonesia," *Sebatik*, vol. 27, no. 1, hlm. 63–69, Jun 2023, doi: 10.46984/sebatik.v27i1.2072.