Jurnal Sistem Komputer dan Informatika (JSON)
Volume 4, Nomor 2, Desember 2022
e-ISSN 2685-998X
DOI 10.30865/json.v4i2.5374

Implementasi Natural Language Processing (NLP) dan Algoritma Cosine Similarity dalam Penilaian Ujian Esai Otomatis

Daniel Oktodeli Sihombing

Program Studi Sistem dan Teknologi Informasi, Institut Teknologi dan Bisnis Sabda Setia, Pontianak, Indonesia
Email: daniel.oktodeli@itbss.ac.id
Email Penulis Korespondensi: daniel.oktodeli@itbss.ac.id
Submitted: **19/12/2022**; Accepted: **31/12/2022**; Published: **31/12/2022**

Abstrak–Evaluasi pembelajaran merupakan suatu kegiatan yang rutin dilaksanakan dalam proses perkuliahan. Ujian esai merupakan tes berupa soal yang bertujuan agar jawaban yang diberikan berupa uraian berdasarkan pemahaman mahasiswa sesuai dengan apa yang diketahuinya. Hasil jawaban yang bervariasi menjadi pertimbangan tersendiri dalam mengoreksi apakah jawaban tersebut sesuai dengan kunci jawaban atau tidak. Hal ini mengakibatkan setiap soal pada ujian esai memiliki bobotnya masing-masing yang nantinya akan dijumlahkan secara kumulatif untuk mendapatkan skor total. Penelitian ini mengimplementasikan Natural Language Processing (NLP) dan algoritma Cosine Similarity untuk melakukan penilaian ujian esai secara otomatis. Document Similarity merupakan salah satu tugas dalam Natural Language Processing (NLP) untuk memeriksa tingkat kemiripan dokumen. Algoritma yang digunakan untuk memeriksa tingkat kemiripan ini adalah Cosine Similarity dimana menggunakan dua buah vektor untuk mengukur tingkat kemiripan dokumen dengan hasil rentang nilai antara 0 sampai dengan 1. Pengolahan data jawaban mahasiswa untuk tiga soal esai mendapatkan hasil yang diharapkan. Hasil perhitungan Cosine Similarity pada soal no 1 menunjukan mahasiswa M3 memiliki jawaban dengan tingkat kemiripan 90,58%. Sedangkan untuk soal no 2 mahasiswa M1 memiliki jawaban dengan tingkat kemiripan 87,71% dan terakhir untuk soal no 3 mahasiswa M1 memiliki jawaban dengan tingkat kemiripan 76,70%.

Kata Kunci: Natural Language Processing; Cosine Similarity; Document Similarity; Penilaian; Esai

Abstract–Evaluation of learning is an activity that is routinely carried out in the lecture process. The essay exam is a test in the form of questions that aim so that the answers given are in the form of descriptions based on student's understanding in accordance with what they know. The results of the various answers are a separate consideration in correcting whether the answer is in accordance with the answer key or not. This resulted in each question on the essay exam having its own weight which would later be added up cumulatively to get a total score. This study implements Natural Language Processing (NLP) and Cosine Similarity algorithms to automatically assess essay exams. Document Similarity is one of the tasks in Natural Language Processing (NLP) to check the degree of document similarity. The algorithm used to check the level of similarity is Cosine Similarity which uses two vectors to measure the degree of similarity of documents with the results ranging from 0 to 1. Processing student answer data for three essay questions gets the expected results. The results of the Cosine Similarity calculation in question no 1 show that M3 students have answers with a similarity level of 90.58%. Whereas for question no 2 M1 students had answers with a similarity level of 87.71% and finally for question no 3 M1 students had answers with a similarity level of 76.70%.

Keywords: Natural Language Processing; Cosine Similarity; Document Similarity; Evaluation; Essay

1. PENDAHULUAN

Evaluasi pembelajaran di dalam Pendidikan Tinggi secara umum terbagi atas dua tahap yaitu evaluasi tengah semester atau yang lebih dikenal dengan Ujian Tengah Semester (UTS) dan evaluasi akhir semester atau yang juga lebih dikenal dengan istilah Ujian Akhir Semester (UAS). Evaluasi ini biasanya menggunakan sistem ujian esai dengan jumlah soal tertentu untuk mengevaluasi sejauh mana pemahaman peserta didik mengenai pembelajaran diperkuliahan yang telah dilaksanakan.

Ujian adalah salah satu alat penialaian dalam proses belajar mengajar yang dilaksanakan secara menyeluruh untuk menentukan kualitas pembelajaran dan merupakan suatu alat ukur untuk mengukur perfoma peserta didik [1]. Ujian esai merupakan ujian dalam bentuk pertanyaan yang bertujuan agar jawaban yang diberikan berupa uraian berdasarkan pemahaman peserta ujian itu sendiri terhadap pertanyaan yang diberikan sesuai dengan apa yang diketahuinya. Jawaban dalam bentuk uraian ini menghasilkan bentuk-bentuk kalimat yang bervariasi sehingga setiap jawaban dari setiap peserta tentu memiliki pola kalimat yang berbeda-beda satu sama lainnya.

Jawaban yang berbeda-beda tersebut akan dikoreksi satu-persatu untuk memahami pola pikir peserta ujian apakah jawaban yang diberikan sesuai dengan kunci jawaban yang sudah ditentukan sebelumnya. Dalam hal ini, mendapatkan jawaban yang persis sama dengan kunci jawaban maupun yang mendekati dengan apa yang dimaknai dalam kunci jawaban menjadi pertimbangan tersendiri bagi dosen yang mengoreksi jawaban tersebut. Jawaban yang memiliki pola kalimat berbeda namun selama makna dari jawaban yang diberikan sama dengan kunci jawaban dari pertanyaan tersebut maka jawaban tersebut tidak salah dan memiliki nilai dengan bobot tertentu.

Semakin banyaknya soal-soal yang diberikan di dalam ujian esai mengakibatkan perlunya waktu yang cukup Panjang untuk mengoreksi jawaban-jawaban tersebut. Permasalahan dalam hal waktu ini menjadi salah satu dasar penelitian ini dilakukan. Serta bagaimana komputer dapat membantu melakukan penilaian ujian esai otomatis menjadi tujuan dari penelitian ini. Penulis melakukan implementasi dasar dari Natural Language Processing (NLP) untuk memproses jawaban dari peserta ujian dan melakukan perhitungan penilaian kimiripan jawaban peserta dengan kunci jawaban menggunakan algoritma Cosine Similarity.

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 4, Nomor 2, Desember 2022
e-ISSN 2685-998X
DOI 10.30865/json.v4i2.5374

Menurut Chowdhury Natural Language Processing (NLP) adalah area penelitian dan aplikasi yang mengeksplorasi bagaimana komputer dapat digunakan untuk memahami dan memanipulasi teks atau ucapan bahasa alami untuk melakukan hal-hal yang bermanfaat [2]. Sedangkan menurut Liddy Natural Language Processing (NLP) adalah pendekatan komputerisasi untuk menganalisis teks yang didasarkan pada seperangkat teori dan seperangkat teknologi [3]. Dengan demikian dapat dipahami juga bahwa Natural Language Processing (NLP) adalah suatu pendekatan terkomputerisasi yang diterapkan untuk memahami dan menganalisis suatu teks atau bahasa alami.

Penelitian sebelumnya yang dilakukan oleh Amalia, dkk yaitu membuat suatu sistem ujian esai online dengan penilaian kemiripan jawaban menggunakan metode Cosine Similarity dan persamaan Term Frequency (TF) untuk menyamakan frekuensi setiap kata yang terdapat dalam kalimat. Untuk pengujian akurasi metode dilakukan pengujian precision, recall, dan f-measure dan berdasarkan hasil analisis dengan menggunakan metode yang telah dicoba diperoleh rata-rata 81% [4]. Selain itu Mufiid, Lestanti dan Kholila melakukan penelitian dengan merancang suatu aplikasi sistem penilaian jawaban esai otomatis menggunakan metode cosine similarity dan synonym recognition. Sistem penilaian jawaban esai otomatis tersebut diuji menggunakan uji akurasi Root Mean Squared Error (RMSE) [1].

Penelitian lain dilakukan oleh Arfandy dan Musdar adalah membuat sistem cerdas Automated Essay Scoring (AES) yang dapat digunakan untuk melakukan penilaian otomatis untuk soal uraian dengan membandingkan jawaban dari peserta didik dengan kunci jawaban dengan menggunakan algoritma Cosine Similarity. Hasil dari penelitian yang dilakukan tersebut menunjukan bahwa algoritma Cosine Similarity berhasil diimplementasikan untuk memberikan nilai ujian esai secara otomatis pada sebuah sistem ujian berbasis komputer [5].

Penelitian lain yang menjadi referensi penulis adalah mengenai penerapan Natural Language Processing (NLP) menggunakan metode Fuzzy String Matching dalam mendeteksi kemiripan artikel melalui keywords. Proses pengujian dilakukan dalam penelitian ini adalah dengan menghitung jumlah kata kunci pada sumber kemudian membandingkan semua kunci sumber dengan artikel tujuan di dalam basis data dengan cara mencari dan membandingkan setiap kata. Dari pengujian data yang dilakukan diperoleh hasil kemiripan yang bervariasi karena kata yang benar sama ditambahkan dengan kata yang mirip dengan kata kunci pada huruf awal dan akhir serta panjang kata tersebut [6].

Penelitian mengenai pengujian terhadap kemiripan kalimat judul tugas akhir yang dilakukan oleh Mawanta, Gunawan dan Wanayumini menjadi referensi selanjutnya yang penulis gunakan dalam penelitian ini. Metode Cosine Similarity dan pembobotan TF-IDF digunakan untuk menguji kemiripan kalimat judul tugas akhir tersebut. Penelitian tersebut memperoleh hasil dimana 43% judul yang diajukan tidak layak untuk diajukan kembali dan 53% layak untuk diajukan sebagai judul tugas akhir. Hal ini dikarenakan judul yang tidak layak tersebut memiliki kemiripan yang tinggi pada judul laporan tugas akhir [7].

Berdasarkan penelitian-penelitian sebelumnya diatas penulis melakukan penelitian mengenai implementasi Natural Language Processing (NLP) menggunakan algoritma Cosine Similarity dalam melakukan penilaian ujian esai otomatis. Penelitian ini difokuskan pada implementasi Natural Language Processing (NLP) mulai dari text preprocessing yang lebih mendalam dan pengecekan document similarity menggunakan algoritma Cosine Similarity untuk melakukan penilaian jawaban esai dengan kunci jawaban dari soal yang diberikan. Hasil penelitian ini diharapkan dapat menjadi solusi dalam melakukan penilaian esai secara otomatis sehingga mempermudah dosen dalam memperoleh nilai hasil evaluasi dengan waktu yang lebih singkat dibandingkan dengan melakukan koreksi jawaban esai secara manual.

2. METODOLOGI PENELITIAN

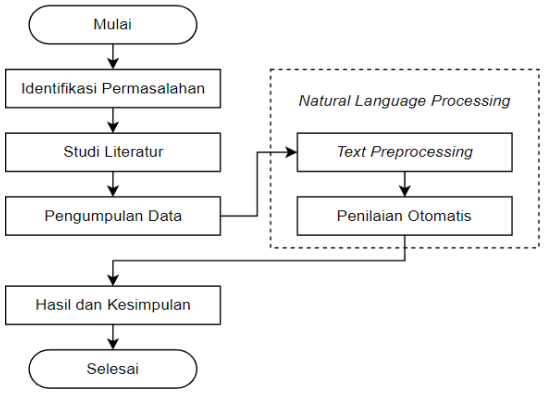
2.1 Tahapan Penelitian

Penelitian ini dilakukan oleh penulis dengan melalui beberapa tahapan atau alur penelitian. Penelitian dimulai dengan melakukan identifikasi permasalahan dimana identifikasi permasalahan penelitian menjadi aspek yang paling penting dalam pelaksanaan penelitian untuk memungkinkan dilakukannya investigasi secara empiris [8]. Setelah permasalahan berhasil diidentifikasi maka tahapan selanjutnya adalah melakukan studi literatur terhadap jurnal-jurnal penelitian terkait dengan topik yang diteliti serta mencari referensi-referensi pendukung lain baik itu melalui e-book maupun website berisikan lliterasi terkait penelitian yang dilakukan, tahapan ini perlu diakukan karena studi literatur merupakan suatu cara untuk menyelesaikan persoalan dengan menelusuri sumber-sumber tulisan yang pernah dibuat sebelumnya [9].

Tahapan selanjutnya penulis melakukan pengumpulan data-data yang akan digunakan untuk melakukan implementasi dari Natural Language Processing (NLP) mulai dari text preprocessing sampai dengan penilaian otomatis. Data soal yang digunakan dalam penelitian ini adalah soal-soal di dalam mata kuliah Struktur Data di Program Studi Sistem dan Teknologi Informasi Institut Teknologi dan Bisnis Sabda Setia dengan kunci jawaban yang telah disusun oleh dosen pengampu mata kuliah tersebut. Setelah Natural Language Processing (NLP) selesai dilakukan maka tahapan selanjutnya adalah merangkum hasil penelitian dan menarik kesimpulan dari hasil penelitian yang telah dilakukan.

Tahapan penelitian yang dilakukan oleh penulis menjadi pedoman untuk melaksanakan penelitian dari awal hingga selesai. Uraian tahapan penelitian yang dilakukan oleh penulis digambarkan sebagai berikut ini.

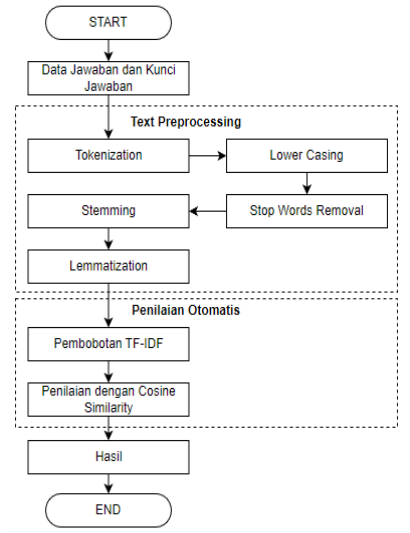
Jurnal Sistem Komputer dan Informatika (JSON)
Volume 4, Nomor 2, Desember 2022
e-ISSN 2685-998X
DOI 10.30865/json.v4i2.5374



Gambar 1. Tahapan Penelitian

2.2 Natural Language Processing (NLP)

Natural Language Processing (NLP) adalah sub-bidang kecerdasan buatan yang berfokus pada interaksi antara komputer dan bahasa manusia. Tujuan dari teknologi ini adalah untuk membuat mesin mampu membaca dan menalar dengan bahasa manusia dan secara otomatis memprosesnya [10]. Pada penelitian ini tahapan yang dilakukan dalam Natural Language Processing (NLP) adalah sebagai berikut.



Gambar 2. Tahapan Natural Language Processing (NLP) Penilaian Esai Otomatis

2.3 Text Preprocessing

Tahapan awal dalam Natural Language Processing (NLP) adalah melakukan Text Preprocessing terlebih dahulu. Text Preprocessing adalah langkah pertama dalam proses membangun model. Langkah-langkah yang digunakan dalam melakukan text preprocessing adalah sebagai berikut [11]:

a. Tokenization
   Memisahkan kalimat menjadi kata-kata.
b. Lower Casing
   Mengubah setiap kata menjadi huruf kecil
c. Stopwords Removal
   Stopwords adalah kata-kata yang sangat umum digunakan dalam dokumen. Kata-kata ini tidak terlalu berarti karena tidak membantu dalam membedakan dua dokumen dan dapat dihilangkan.
d. Stemming
   Proses mengubah kata kebentuk akarnya
e. Lemmatization
   Tidak seperti stemming, lemmatization mengurangi kata menjadi hanya kata yang ada dalam bahasa Indonesia saja

2.4 TF (Term Frequency)

Term Frequency atau TF adalah jumlah kuantitas term yang sering muncul dari dalam suatu dokumen atau teks. Proses TF ini menghitung jumlah kemunculan (frekuensi) term t_i dalam setiap dokumen d_j [4].

$$W_{TF}(t_i, d_j) = f(t_i, d_j) \quad (1)$$

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 4, Nomor 2, Desember 2022
e-ISSN 2685-998X
DOI 10.30865/json.v4i2.5374

2.5 Cosine Similarity

Cosine similarity atau cosim merupakan algoritma yang digunakan untuk mengukur tingkat kemiripan nilai sudut cosinus dari perkalian dua vektor yang dibandingkan, vektor tersebut dinyatakan mirip apabila sudut cosim bernilai 1, dimana nilai cosinus 0˚adalah 1 dan memiliki nilai kurang dari 1 untuk nilai sudut yang lainnya [12]. Perhitungan dalam Cosine Similarity ditunjukan pada persamaan (4)

$$\text{Similarity} = \text{Cos}(\theta) = \frac{A.B}{\|A\|\|B\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} (A_i)^2}\sqrt{\sum_{i=1}^{n} (B_i)^2}} \quad (2)$$

3. HASIL DAN PEMBAHASAN

Implementasi Natural Language Processing (NLP) dimulai dari tahapan persiapan data yang akan diproses. Data tersebut merupakan kunci jawaban dari soal esai mata kuliah struktur data beserta sampel data jawaban dari mahasiswa sebagai data awal untuk menguji ketepatan penilaian esai otomatis menggunakan algoritma Cosine Similarity. Berikut ini adalah tabel data untuk soal esai mata kuliah struktur data beserta kunci jawabannya.

Tabel 1. Soal dan Kunci Jawaban Esai

No	Soal	Kunci Jawaban (Q)	Bobot
1	Menurut anda apa yang dimaksud dengan struktur data?	Struktur data adalah suatu cara pengelolaan data mulai dari penyimpanan, pengorganisasian dan penyimpanan data di dalam media penyimpanan komputer agar dapat digunakan secara efisien	35
2	Apa perbedaan antara tipe data integer dan float?	Tipe data integer digunakan untuk merepresentasikan suatu bilangan bulat positif maupun negatif sedangkan tipe data float digunakan untuk merepresentasikan bilangan pecahan positif maupun negatif	35
3	Jelaskan secara ringkas karakteristik array yang anda ketahui!	Array memiliki karakteristik berupa kumpulan data dengan tipe data yang sama atau bersifat homogen, memiliki dimensi 1, 2 dan 3 atau multidimensi serta dapat diakses secara acak	30
		Total	100

Data sampel yang digunakan adalah 5 data jawaban mahasiswa diambil dari data-data jawaban mahasiswa yang telah mengerjakan soal esai tersebut. Berikut ini adalah tabel data jawaban mahasiswa yang akan dilakukan proses penilaian otomatis menggunakan training mengenai sampel jawaban yang digunakan untuk penilaian esai otomatis adalah sebagai berikut.

Tabel 2. Data Jawaban Mahasiswa Soal No 1

Mahasiswa	Jawaban Soal No 1
M1	Cara untuk mengatur, mengolah, menyimpan data secara terstruktur agar lebih mudah untuk diakses di penyimpanan komputer
M2	Struktur data adalah suatu cara menyimpan, mengatur, dan mengelompokkan data di dalam suatu memori komputer agar dapat terhubung satu sama lain dan bisa digunakan secara efisien.
M3	Struktur data sendiri merupakan cara pengorganisasian, penyimpanan dan pengaturan suatu data dalam penyimpanan komputer agar dapat digunakan secara efisien.
M4	Suatu cara untuk menyimpan, mengatur, mengelola data-data secara terstruktur pada sistem komputer agar menjadi hal yang berguna dan mudah diakses.
M5	Struktur data adalah suatu susunan data yang saling terhubung satu sama lain dan teratur, yang tersimpan di dalam memori komputer.

Tabel 3. Data Jawaban Mahasiswa Soal No 2

Mahasiswa	Jawaban Soal No 2
M1	Integer adalah tipe data yang terdiri dari bilangan bulat positif maupun negatif, sedangkan float berisi bilangan pecahan positif dan negatif. (Sama-sama Primitive structure data)
M2	Kalau integer itu untuk menyimpan data bilangan bulat kalau float itu untuk menyimpan data bilangan desimal.
M3	Perbedaan dari tipe data integer dan float tentu pada nilai nya yang mana integer merupakan nilai bilangan bulan yang tidak memiliki pecahan atau desimal, sedangkan float sendiri merupakan tipe data bilangan pecahan atau desimal.

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 4, Nomor 2, Desember 2022
e-ISSN 2685-998X
DOI 10.30865/json.v4i2.5374

Mahasiswa	Jawaban Soal No 2
M4	Integer adalah bilangan bulat (bukan pecahan), sedangkan float dapat terdiri dari bilangan koma (pecahan desimal).
M5	Integer adalah tipe data yang menyimpan bilangan data bulat sedangkan float adalah tipe data yang mampu menyimpan data desimal.

Tabel 4. Data Jawaban Mahasiswa Soal No 3

Mahasiswa	Jawaban Soal No 3
M1	Array adalah sekumpulan data yang memiliki tipe data yang sama dan dapat diakses untuk keperluan tertentu, array bisa berdimensi 1, 2, 3.
M2	Array itu menyimpan data yang tipenya sama. Cara untuk inisialisasinya adalah tipe data nama array [jumlah elemen]. Array selalu ada []. Indeks dimulai dari nol.
M3	Array sendiri memiliki karakteristik berupa terdiri dari kumpulan data yang homogen atau sama dengan Inisialisasi satu nama, dan array juga terbagi menjadi satu dimensi, dua dimensi dan multidimensi, dan array juga memiliki index yang dimulai dari 0
M4	Array diakses berdasarkan indeks sehingga memudahkan dalam penggunaan data, Array menyimpan beberapa nilai dari tipe data yang sama
M5	Array mampu menyimpan tipe data yang sama dalam satu susunan. Tiap variabel tersimpan di index yang berbeda-beda. index array dimulai dari 0.

3.1 Text Preprocessing

Setelah memperoleh data jawaban dari mahasiswa mengenai soal esai yang diberikan, maka tahapan selanjutnya adalah melakukan text preprocessing terhadap kunci jawaban soal dan data jawaban tersebut. Text preprocessing merupakan tahapan awal dalam metode NLP untuk mempersiapkan teks yang tidak terstruktur menjadi data yang baik dan siap untuk diolah [13].

3.1.1 Tokenization, Lower Casing dan Stopwords Removal

Tokenization merupakan suatu proses untuk memecah teks dokumen menjadi kata serta menghilangkan angka, tanda baca dan spasi [14]. Lower Casing merupakan proses untuk mengubah huruf dalam kata yang digunakan seluruhnya menjadi huruf kecil. Sedangkan Stopwords Removal merupakan proses selanjutnya untuk menghilangkan kata-kata stop-word yang dapat dihilangkan untuk memperoleh hasil yang lebih akurat. Tahap stopwords removal dilakukan berdasarkan kamus stop-word yang diperoleh dari jurnal penelitian oleh Fadillah Z Tala [15]. Pada tahapan ini kunci jawaban dan data jawaban mahasiswa dilakukan proses tokenization, lower casing dan stopwords removal hingga menghasilkan data seperti pada tabel 5 berikut ini.

Tabel 5. Hasil Tokenization, Lower Casing dan Stopwords Removal Kunci (Q) dan Data Jawaban (M) Soal No 1

Teks	Hasil Tokenization, Lower Casing dan Stopwords Removal
Q1	struktur data pengelolan data penyimpanan pengorganisasian penyimpanan data media penyimpanan komputer efisien
M1	mengatur mengolah menyimpan data terstruktur mudah diakses penyimpanan komputer
M2	struktur data menyimpan mengatur mengelompokkan data memori komputer terhubung efisien
M3	struktur data pengorganisasian penyimpanan pengaturan data penyimpanan komputer efisien
M4	menyimpan mengatur mengelolah datadata terstruktur sistem komputer berguna mudah diakses
M5	struktur data susunan data terhubung teratur tersimpan memori komputer

Selanjutnya lakukan proses tokenization, lower casing dan stopwords removal yang sama untuk soal no 2 dan no 3. Semakin banyak soal ujian esai maka proses tersebut dilakukan sebanyak jumlah soal yang ada digunakan. Hasil dari tahapan ini kemudian dilanjutkan kepada proses Stemming dan Lemmatization.

3.1.2 Stemming dan Lemmatization

Stemming adalah proses dalam mentransformasi suatu kata menjadi kata dasarnya dengan menghilangkan imbuhan dari kata tersebut sehingga menghasilkan dimensi kata yang berbeda namun memiliki makna yang sama [15]. sedangkan lemmatization merupakan proses menemukan bentuk dasar sebuah kata yang memiliki arti tertentu berdasarkan kamus [16]. Hasil proses stemming dan lemmatization dari kunci jawaban dan data jawaban ditunjukan pada tabel berikut ini.

Tabel 6. Hasil Proses Stemming dan Lemmatization Kunci (Q) dan Data Jawaban (M) Soal No 1

Teks	Hasil Proses Stemming dan Lemmatization
Q1	struktur data kelola data simpan organisasi simpan data media simpan komputer efisien
M1	atur olah simpan data struktur mudah akses simpan komputer
M2	struktur data simpan atur kelompok data memori komputer hubung efisien
M3	struktur data organisasi simpan atur data simpan komputer efisien

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 4, Nomor 2, Desember 2022
e-ISSN 2685-998X
DOI 10.30865/json.v4i2.5374

Teks	Hasil Proses Stemming dan Lemmatization
M4	simpan atur kelola data data struktur sistem komputer guna mudah akses
M5	struktur data susun data hubung atur simpan memori komputer

Tabel 7. Hasil Proses Stemming dan Lemmatization Kunci (Q) dan Data Jawaban (M) Soal No 2

Teks	Hasil Proses Stemming dan Lemmatization
Q2	tipe data representasi bilang bulat positif negatif tipe data representasi bilang pecah positif negatif
M1	tipe data bilang bulat positif negatif isi bilang pecah positif negatif data
M2	simpan data bilang bulat simpan data bilang desimal
M3	beda tipe data nilai nilai bilang milik pecah desimal tipe data bilang pecah desimal
M4	bilang bulat pecah bilang koma pecah desimal
M5	tipe data simpan bilang data bulat tipe data simpan data desimal

Tabel 8. Hasil Proses Stemming dan Lemmatization Kunci (Q) dan Data Jawaban (M) Soal No 3

Teks	Hasil Proses Stemming dan Lemmatization
Q3	milik karakteristik kumpul data tipe data sifat homogen milik dimensi multidimensi akses acak
M1	kumpul data milik tipe data akses perlu mens
M2	simpan data tipe inisial tipe data nama elemen indeks nol
M3	milik karakteristik kumpul data homogen inisial nama bagi dimensi dimensi multidimensi milik
M4	akses dasar indeks mudah guna data simpan nilai tipe data
M5	simpan tipe data susun variabel simpan beda beda

3.2 Penilaian Otomatis

Setelah melalui tahap text preprocessing kemudian dilanjutkan dengan memulai tahap penilaian otomatis yaitu dengan memberikan nilai pembobotan terlebih dahulu menggunakan TF (Term Frequency) untuk menghasilkan jumlah frekuensi kata dalam suatu jawaban atau dokumen yang kemudian diproses menggunakan algoritma cosine similarity untuk memberikan nilai data jawaban dengan kunci jawaban. Hasil cosine similarity yang mendekati nilai 1 menunjukan bahwa data jawaban dapat dikatakan mendekati jawaban yang tepat untuk kemudian hasil dari nilai tersebut dikalikan dengan bobot nilai masing-masing soal. Hasil dari perkalian nilai cosine similarity dengan bobot dari masing-masing soal kemudian dijumlahkan untuk memperoleh nilai total dari jawaban mahasiswa tersebut.

3.2.1 TF (Term Frequency)

Perhitungan term frequency dilakukan dengan cara menggabungkan seluruh term jawaban yang unik untuk kemudian dihitung frekuensi setiap kata sebagai nilai bobot TF dari masing-masing term. Berikut ini adalah hasil pembobotan TF dari hasil text preprocessing kunci jawaban dan data jawaban mahasiswa.

Tabel 9. Hasil Pembobotan TF dari Kunci Jawaban (Q) dan Data Jawaban (D) Soal No 1

Term (t)	Q1	D1	D2	D3	D4	D5
struktur	1	1	1	1	1	1
data	3	1	2	2	2	2
kelola	1	0	0	0	1	0
simpan	3	2	1	2	1	1
organisasi	1	0	0	1	0	0
media	1	0	0	0	0	0
komputer	1	1	1	1	1	1
efisien	1	0	1	1	0	0
atur	0	1	1	1	1	1
olah	0	1	0	0	0	0
mudah	0	1	0	0	1	0
akses	0	1	0	0	1	0
kelompok	0	0	1	0	0	0
memori	0	0	1	0	0	1
hubung	0	0	1	0	0	1
sistem	0	0	0	0	1	0
guna	0	0	0	0	1	0
susun	0	0	0	0	0	1

Tabel 10. Hasil Pembobotan TF dari Kunci Jawaban (Q) dan Data Jawaban (D) Soal No 2

Term (t)	Q2	D1	D2	D3	D4	D5
tipe	2	1	0	2	0	2
data	2	2	2	2	0	4

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 4, Nomor 2, Desember 2022
e-ISSN 2685-998X
DOI 10.30865/json.v4i2.5374

Term (t)	Q2	D1	D2	D3	D4	D5
representasi	2	0	0	0	0	0
bilang	2	2	2	2	2	1
bulat	1	1	1	0	1	1
positif	2	2	0	0	0	0
negatif	2	2	0	0	0	0
pecah	1	1	0	2	2	0
isi	0	1	0	0	0	0
simpan	0	0	2	0	0	2
desimal	0	0	1	2	1	1
beda	0	0	0	1	0	0
nilai	0	0	0	2	0	0
milik	0	0	0	1	0	0
koma	0	0	0	0	1	0

Tabel 11. Hasil Pembobotan TF dari Kunci Jawaban (Q) dan Data Jawaban (D) Soal No 3

Term (t)	Q3	D1	D2	D3	D4	D5
milik	2	1	0	2	0	0
karakteristik	1	0	0	1	0	0
kumpul	1	1	0	1	0	0
data	2	2	2	1	2	1
tipe	1	1	2	0	1	1
sifat	1	0	0	0	0	0
homogen	1	0	0	1	0	0
dimensi	1	1	0	2	0	0
multidimensi	1	0	0	1	0	0
akses	1	1	0	0	1	0
acak	1	0	0	0	0	0
perlu	0	1	0	0	0	0
simpan	0	0	1	0	1	2
inisial	0	0	1	1	0	0
nama	0	0	1	1	0	0
elemen	0	0	1	0	0	0
indeks	0	0	1	0	1	0
nol	0	0	1	0	0	0
bagi	0	0	0	1	0	0
dasar	0	0	0	0	1	0
mudah	0	0	0	0	1	0
guna	0	0	0	0	1	0
nilai	0	0	0	0	1	0
susun	0	0	0	0	0	1
variabel	0	0	0	0	0	1
beda	0	0	0	0	0	2

3.2.2 Cosine Similarity

Setelah melakukan pembobotan TF terhadap kunci jawaban dan data jawaban dari masing-masing soal, berikutnya dilakukan tahap pengecekan similaritas antara dokumen jawaban dan kunci jawaban. Pengecekan similaritas ini menggunakan algoritma Cosine Similarity. Cosine similarity merupakan algoritma yang digunakan untuk menghitung tingkat kesamaan atau similarity antar dua buah objek atau dokumen yang dinyatakan dalam dua buah vektor dimana tingkat similarity yang dihasilkan berada dalam kisaran interval 0 (nol) sampai dengan 1 (satu). Nilai 0 (nol) menandakan kedua dokumen tersebut sama sekali berbeda, sedangkan nilai 1 (satu) menandakan bahwa kedua dokumen tersebut persis sama [17]. Berikut ini adalah hasil perhitungan nilai Cosine Similarity dari masing-masing kunci jawaban dan data jawaban tiap soal.

Tabel 12. Hasil Pembobotan Cosine Similarity Soal No 1

Terms (t)	Q1.D1	Q1.D2	Q1.D3	Q1.D4	Q1.D5	$(Q1)^2$	$(D1)^2$	$(D2)^2$	$(D3)^2$	$(D4)^2$	$(D5)^2$
struktur	1	1	1	1	1	1	1	1	1	1	1
data	3	6	6	6	6	9	1	4	4	4	4
kelola	0	0	0	1	0	1	0	0	0	1	0
simpan	6	3	6	3	3	9	4	1	4	1	1
organisasi	0	0	1	0	0	1	0	0	1	0	0

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 4, Nomor 2, Desember 2022
e-ISSN 2685-998X
DOI 10.30865/json.v4i2.5374

Terms (t)	Q1.D1	Q1.D2	Q1.D3	Q1.D4	Q1.D5	$(Q1)^2$	$(D1)^2$	$(D2)^2$	$(D3)^2$	$(D4)^2$	$(D5)^2$
media	0	0	0	0	0	1	0	0	0	0	0
komputer	1	1	1	1	1	1	1	1	1	1	1
efisien	0	1	1	0	0	1	0	1	1	0	0
atur	0	0	0	0	0	0	1	1	1	1	1
olah	0	0	0	0	0	0	1	0	0	0	0
mudah	0	0	0	0	0	0	1	0	0	1	0
akses	0	0	0	0	0	0	1	0	0	1	0
kelompok	0	0	0	0	0	0	0	1	0	0	0
memori	0	0	0	0	0	0	0	1	0	0	1
hubung	0	0	0	0	0	0	0	1	0	0	1
sistem	0	0	0	0	0	0	0	0	0	1	0
guna	0	0	0	0	0	0	0	0	0	1	0
susun	0	0	0	0	0	0	0	0	0	0	1
Total (Σ)	**11**	**12**	**16**	**12**	**11**	**24**	**11**	**12**	**13**	**13**	**11**

Tabel 13. Hasil Perhitungan Nilai Cosine Similarity Soal No 1

Mahasiswa (M)	$\Sigma(Q.D)$	$\Sigma(Q)^2$	$\Sigma(D)^2$	$\sqrt{\Sigma(Q)^2}$	$\sqrt{\Sigma(D)^2}$	$\sqrt{\Sigma(Q)^2}.\sqrt{\Sigma(D)^2}$	Cosine Similarity
M1	11	24	11	4,8989794856	3,3166247904	16,2480768093	0,67700320
M2	12	24	12	4,8989794856	3,4641016151	16,9705627485	0,70710678
M3	16	24	13	4,8989794856	3,6055512755	17,6635217327	0,90582163
M4	12	24	13	4,8989794856	3,6055512755	17,6635217327	0,67936622
M5	11	24	11	4,8989794856	3,3166247904	16,2480768093	0,67700320

Tabel 14. Hasil Pembobotan Cosine Similarity Soal No 2

Terms (t)	Q1.D1	Q1.D2	Q1.D3	Q1.D4	Q1.D5	$(Q1)^2$	$(D1)^2$	$(D2)^2$	$(D3)^2$	$(D4)^2$	$(D5)^2$
tipe	2	0	4	0	4	4	1	0	4	0	4
data	4	4	4	0	8	4	4	4	4	0	16
representasi	0	0	0	0	0	4	0	0	0	0	0
bilang	4	4	4	4	2	4	4	4	4	4	1
bulat	1	1	0	1	1	1	1	1	0	1	1
positif	4	0	0	0	0	4	4	0	0	0	0
negatif	4	0	0	0	0	4	4	0	0	0	0
pecah	1	0	2	2	0	1	1	0	4	4	0
isi	0	0	0	0	0	0	1	0	0	0	0
simpan	0	0	0	0	0	0	0	4	0	0	4
desimal	0	0	0	0	0	0	0	1	4	1	1
beda	0	0	0	0	0	0	0	0	1	0	0
nilai	0	0	0	0	0	0	0	0	4	0	0
milik	0	0	0	0	0	0	0	0	1	0	0
koma	0	0	0	0	0	0	0	0	0	1	0
Total (Σ)	**20**	**9**	**14**	**7**	**15**	**26**	**20**	**14**	**26**	**11**	**27**

Tabel 15. Hasil Perhitungan Nilai Cosine Similarity Soal No 2

Mahasiswa (M)	$\Sigma(Q.D)$	$\Sigma(Q)^2$	$\Sigma(D)^2$	$\sqrt{\Sigma(Q)^2}$	$\sqrt{\Sigma(D)^2}$	$\sqrt{\Sigma(Q)^2}.\sqrt{\Sigma(D)^2}$	Cosine Similarity
M1	20	26	20	5,0990195136	4,4721359550	22,803508502	0,87705802
M2	9	26	14	5,0990195136	3,7416573868	19,078784028	0,47172818
M3	14	26	26	5,0990195136	5,0990195136	26,000000000	0,53846154
M4	7	26	11	5,0990195136	3,3166247904	16,911534525	0,41391868
M5	15	26	27	5,0990195136	5,1961524227	26,495282599	0,56613852

Tabel 16. Hasil Pembobotan Cosine Similarity Soal No 3

Terms (t)	Q1.D1	Q1.D2	Q1.D3	Q1.D4	Q1.D5	$(Q1)^2$	$(D1)^2$	$(D2)^2$	$(D3)^2$	$(D4)^2$	$(D5)^2$
milik	2	0	4	0	0	4	1	0	4	0	0
karakteristik	0	0	1	0	0	1	0	0	1	0	0
kumpul	1	0	1	0	0	1	1	0	1	0	0
data	4	4	2	4	2	4	4	4	1	4	1
tipe	1	2	0	1	1	1	1	4	0	1	1
sifat	0	0	0	0	0	1	0	0	0	0	0

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 4, Nomor 2, Desember 2022
e-ISSN 2685-998X
DOI 10.30865/json.v4i2.5374

Terms (t)	Q1.D1	Q1.D2	Q1.D3	Q1.D4	Q1.D5	$(Q1)^2$	$(D1)^2$	$(D2)^2$	$(D3)^2$	$(D4)^2$	$(D5)^2$
homogen	0	0	1	0	0	1	0	0	1	0	0
dimensi	1	0	2	0	0	1	1	0	4	0	0
multidimensi	0	0	1	0	0	1	0	0	1	0	0
akses	1	0	0	1	0	1	1	0	0	1	0
acak	0	0	0	0	0	1	0	0	0	0	0
perlu	0	0	0	0	0	0	1	0	0	0	0
simpan	0	0	0	0	0	0	0	1	0	1	4
inisial	0	0	0	0	0	0	0	1	1	0	0
nama	0	0	0	0	0	0	0	1	1	0	0
elemen	0	0	0	0	0	0	0	1	0	0	0
indeks	0	0	0	0	0	0	0	1	0	1	0
nol	0	0	0	0	0	0	0	1	0	0	0
bagi	0	0	0	0	0	0	0	0	1	0	0
dasar	0	0	0	0	0	0	0	0	0	1	0
mudah	0	0	0	0	0	0	0	0	0	1	0
guna	0	0	0	0	0	0	0	0	0	1	0
nilai	0	0	0	0	0	0	0	0	0	1	0
susun	0	0	0	0	0	0	0	0	0	0	1
variabel	0	0	0	0	0	0	0	0	0	0	1
beda	0	0	0	0	0	0	0	0	0	0	4
Total (Σ)	**10**	**6**	**12**	**6**	**3**	**17**	**10**	**14**	**16**	**12**	**12**

Tabel 17. Hasil Perhitungan Nilai Cosine Similarity Soal No 3

Mahasiswa (M)	$\Sigma(Q.D)$	$\Sigma(Q)^2$	$\Sigma(D)^2$	$\sqrt{\Sigma(Q)^2}$	$\sqrt{\Sigma(D)^2}$	$\sqrt{\Sigma(Q)^2}.\sqrt{\Sigma(D)^2}$	Cosine Similarity
M1	10	17	10	4,1231056256	3,1622776602	13,038404810	0,76696499
M2	6	17	14	4,1231056256	3,7416573868	15,427248621	0,38892223
M3	12	17	16	4,1231056256	4,0000000000	16,492422502	0,72760688
M4	6	17	12	4,1231056256	3,4641016151	14,282856857	0,42008403
M5	3	17	12	4,1231056256	3,4641016151	14,282856857	0,21004201

Tabel-tabel diatas merupakan hasil dari tahapan perhitungan nilai Cosine Similarity mulai dari soal no 1 hingga soal no 3 berdasarkan data jawaban dari masing-masing mahasiswa. Setelah nilai Cosine Similarity diperoleh maka proses terakhir yang dilakukan untuk mendapatkan hasil penilaian adalah dengan mengalikan nilai Cosine Similarity dan bobot nilai masing-masing soal yang telah ditentukan sebelumnya.

3.2.3 Hasil Penilaian

Setelah hasil perhitungan Cosine Similarity dari setiap soal diperoleh maka nilai Cosine Similarity tersebut dikalikan dengan bobot nilai setiap soal, bobot nilai untuk masing-masing soal dapat dilihat kembali pada tabel 1. Hasil perkalian tersebut dijumlahkan secara total untuk memperoleh nilai akhir dari masing-masing mahasiswa. Hasil penilaian akhir yang diperoleh dari hasil pengolahan data dengan Natural Language Processing (NLP) dan Algoritma Cosine Similarity ditunjukan pada tabel berikut ini.

Tabel 18. Hasil Tingkat Kemiripan Jawaban Menggunakan NLP dan Cosine Similarity

Mahasiswa (M)	Cosim Soal 1	Cosim Soal 2	Cosim Soal 3	Similarity Q1 (%)	Similarity Q2 (%)	Similarity Q3 (%)
M1	0,6770032	0,87705802	0,76696499	67,70	87,71	76,70
M2	0,70710678	0,47172818	0,38892223	70,71	47,17	38,89
M3	0,90582163	0,53846154	0,72760688	90,58	53,85	72,76
M4	0,67936622	0,41391868	0,42008403	67,94	41,39	42,01
M5	0,6770032	0,56613852	0,21004201	67,70	56,61	21,00

Tabel 19. Hasil Penilaian Esai Otomatis Menggunakan NLP dan Cosine Similarity

Mahasiswa (M)	Cosim Soal 1	Cosim Soal 2	Cosim Soal 3	N1	N2	N3	Nilai = N1+N2+N3
M1	0,67700320	0,87705802	0,76696499	23,70	30,70	23,01	77,41
M2	0,70710678	0,47172818	0,38892223	24,75	16,51	11,67	52,93
M3	0,90582163	0,53846154	0,72760688	31,70	18,85	21,83	72,38
M4	0,67936622	0,41391868	0,42008403	23,78	14,49	12,60	50,87
M5	0,67700320	0,56613852	0,21004201	23,70	19,81	6,30	49,81

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 4, Nomor 2, Desember 2022
e-ISSN 2685-998X
DOI 10.30865/json.v4i2.5374

Data pada tabel diatas menunjukan bahwa hasil perhitungan nilai Cosine Similarity menunjukan dapat digunakan untuk memberikan penilaian otomatis terhadap soal-soal esai yang dijawab oleh mahasiswa. Dengan penilaian multisoal, pembobotan nilai masing-masing soal menjadi dasar perhitungan nilai akhir dari masing-masing soal dan kemudian dijumlahkan seluruhnya untuk memperoleh nilai total hasil ujian esai tiap-tiap mahasiswa. Hasil perhitungan Cosine Similarity pada soal no 1 menunjukan mahasiswa M3 memiliki jawaban dengan tingkat kemiripan 90,58%. Sedangkan untuk soal no 2 mahasiswa M1 memiliki jawaban dengan tingkat kemiripan 87,71% dan terakhir untuk soal no 3 mahasiswa M1 memiliki jawaban dengan tingkat kemiripan 76,70%.

4. KESIMPULAN

Soal esai memiliki tingkat kesulitan tersendiri dalam memberikan penilaian terhadap jawaban-jawaban yang diperoleh dari hasil kerja mahasiswa. Banyaknya kemungkinan pola kalimat yang digunakan oleh mahasiswa dalam menuangkan pola pikirnya untuk menjawab pertanyaan setiap soal menjadi hal yang perlu diperhatikan agar penilaian menjadi lebih optimal. Namun kondisi demikian mengakibatkan perlunya waktu yang cukup panjang untuk memeriksanya secara menyeluruh. Penelitian ini melakukan implementasi Natural Language Processing (NLP) dalam melakukan pemrosesan data jawaban mahasiswa. NLP memungkinkan kita untuk mengolah teks dan melakukan pengecekan document similarity. Pengecekan document similarity tersebut digunakan dalam penelitian ini untuk menentukan tingkat kemiripan antara kunci jawaban soal dengan data jawaban mahasiswa. Algoritma yang digunakan untuk melakukan pengecekan document similarity adalah Cosine Similarity. Cosine Similarity menggunakan dua vektor untuk mengukur tingkat kemiripan dokumen dengan rentang nilai antara nol (0) sampai dengan satu (1). Pengolahan terhadap data jawaban dari mahasiswa untuk 3 soal esai memperoleh hasil yang diharapkan. Hasil perhitungan Cosine Similarity pada soal no 1 menunjukan mahasiswa M3 memiliki jawaban dengan tingkat kemiripan 90,58%. Sedangkan untuk soal no 2 mahasiswa M1 memiliki jawaban dengan tingkat kemiripan 87,71% dan terakhir untuk soal no 3 mahasiswa M1 memiliki jawaban dengan tingkat kemiripan 76,70%. Berdasarkan hasil tersebut implementasi Natural Language Processing (NLP) dan algoritma Cosine Similarity dapat digunakan untuk melakukan penilaian ujian esai secara otomatis.

REFERENCES

[1] I. Mufiid, S. Lestanti, and N. Kholila, “APLIKASI PENILAIAN JAWABAN ESAI OTOMATIS MENGGUNAKAN METODE SYNONYM RECOGNITION DAN COSINE SIMILARITY BERBASIS WEB,” Jurnal Mnemonic, vol. 4, no. 2, pp. 31–37, Sep. 2021, doi: 10.36040/mnemonic.v4i2.4067.
[2] G. G. Chowdhury, “Natural language processing,” Annual Review of Information Science and Technology, vol. 37, no. 1, pp. 51–89, Jan. 2005, doi: 10.1002/aris.1440370103.
[3] E. D. Liddy, “Natural Language Processing,” in Encyclopedia of Library and Information Science, 2nd Ed, New York: Marcel Decker, Inc, 2001.
[4] E. L. Amalia, A. J. Jumadi, I. A. Mashudi, and D. W. Wibowo, “Analisis Metode Cosine Similarity Pada Aplikasi Ujian Online Otomatis (Studi Kasus JTI POLINEMA),” Jurnal Teknologi Informasi dan Ilmu Komputer, vol. 8, no. 2, p. 343, Mar. 2021, doi: 10.25126/jtiik.2021824356.
[5] H. Arfandy and I. A. Musdar, “Rancang Bangun Sistem Cerdas Pemberian Nilai Otomatis Untuk Ujian Essai Menggunakan Algoritma Cosine Similarity,” Inspiration: Jurnal Teknologi Informasi dan Komunikasi, vol. 10, no. 2, p. 123, Dec. 2020, doi: 10.35585/inspir.v10i2.2580.
[6] H. Rumapea, “DETEKSI KEMIRIPAN ARTIKEL MELALUI KEYWORDS DENGAN METODE FUZZY STRING MATCHING DALAM NATURAL LANGUAGE PROCESSING,” METHOMIKA Jurnal Manajemen Informatika dan Komputerisasi Akuntansi, vol. 5, no. 1, pp. 60–66, Apr. 2021, doi: 10.46880/jmika.Vol5No1.pp60-66.
[7] I. Mawanta, T. S. Gunawan, and W. Wanayumini, “Uji Kemiripan Kalimat Judul Tugas Akhir dengan Metode Cosine Similarity dan Pembobotan TF-IDF,” JURNAL MEDIA INFORMATIKA BUDIDARMA, vol. 5, no. 2, p. 726, Apr. 2021, doi: 10.30865/mib.v5i2.2935.
[8] A. R. S. Nasution, “Identifikasi Permasalahan Penelitian,” ALACRITY : Journal of Education, pp. 13–19, Jul. 2021, doi: 10.52121/alacrity.v1i2.21.
[9] Salmaa, “Studi Literatur: Pengertian, Ciri-Ciri, dan Teknik Pengumpulan Datanya,” https://penerbitdeepublish.com/studi-literatur/, Jun. 07, 2021.
[10] E. Commission et al., Natural language processing for public services. Publications Office of the European Union, 2022. doi: doi/10.2799/304724.
[11] Harshith, “Text Preprocessing in natural language processing using Python,” https://towardsdatascience.com/text-preprocessing-in-natural-language-processing-using-python-6113ff5decd8, Nov. 21, 2019.
[12] A. R. Lahitani, “Automated Essay Scoring menggunakan Cosine Similarity pada Penilaian Esai Multi Soal,” Jurnal Kajian Ilmiah, vol. 22, no. 2, pp. 107–118, May 2022, doi: 10.31599/jki.v22i2.1121.
[13] Rianita Giovanni Katryn, “Text Preprocessing: Tahap Awal dalam Natural Language Processing (NLP),” https://medium.com/mandiri-engineering/text-preprocessing-tahap-awal-dalam-natural-language-processing-nlp-bc5fbb6606a, Sep. 28, 2020.
[14] P. M. Prihatini, “IMPLEMENTASI EKSTRAKSI FITUR PADA PENGOLAHAN DOKUMEN BERBAHASA INDONESIA,” JURNAL MATRIX, vol. 6, no. 3, pp. 174–178, Nov. 2016, [Online]. Available: http://ojs.pnb.ac.id/index.php/matrix/article/view/151
[15] S. Shevira, I. M. A. D. Suarjaya, and P. W. Buana, “Pengaruh Kombinasi dan Urutan Pre-Processing padaTweets Bahasa Indonesia,” JITTER - Jurnal Ilmiah Teknologi dan Komputer, vol. 3, no. 2, Aug. 2022, [Online]. Available: https://ojs.unud.ac.id/index.php/jitter/article/view/88613

Jurnal Sistem Komputer dan Informatika (JSON)
Volume 4, Nomor 2, Desember 2022
e-ISSN 2685-998X
DOI 10.30865/json.v4i2.5374

[16] Y. Miftahuddin, J. Pardede, and R. Dewi, "Penerapan Algoritma Lemmatization pada Dokumen Bahasa Indonesia," MIND Journal, vol. 3, no. 2, pp. 47–56, Feb. 2019, doi: 10.26760/mindjournal.v3i2.47-56.
[17] A. Apriani, H. Zakiyudin, and K. Marzuki, "Penerapan Algoritma Cosine Similarity dan Pembobotan TF-IDF System Penerimaan Mahasiswa Baru pada Kampus Swasta," Jurnal Bumigora Information Technology (BITe), vol. 3, no. 1, pp. 19–27, Jul. 2021, doi: 10.30812/bite.v3i1.1110.