

Implementasi Algoritma Cosine Similarity Berbasis Natural Language Processing Pada Penilaian Esai Otomatis

Naya Sabilla Putri¹, Alan Novi Tompunu², Adi Sutrisman³

^{1,2,3}Jurusan Teknik Komputer, Teknologi Informatika Multimedia Digital, Politeknik Negeri Sriwijaya, Palembang, Indonesia

Email: ^{1,*}nayasabilla2@gmail.com, ²alan_nt@polsri.ac.id, ³adisutrisman@polsri.ac.id

^{*)} Email Penulis Utama

Abstrak– Evaluasi tes esai berbasis kertas di SMP Srijaya Negara menghadapi kendala durasi koreksi yang lama, risiko salah hitung, serta bias subjektivitas. Demi mengatasi persoalan tersebut, dikembangkan sistem *Computer Based Test* (CBT) berbasis web menggunakan teknologi *Natural Language Processing* (NLP) dan algoritma Cosine Similarity untuk menilai esai otomatis secara objektif. Pengembangan aplikasi ini menerapkan metodologi *Rapid Application Development* (RAD) dengan kombinasi pemrograman PHP, *framework* Laravel, dan *database* MySQL. Dalam prosesnya, modul NLP mengeksekusi empat tahapan *preprocessing*: *case folding*, *tokenizing*, *filtering*, dan *stemming*. Dokumen hasil pembersihan kemudian dikonversi menjadi vektor numerik melalui skema *Bag of Words* (BoW) untuk dihitung skor kemiripannya berdasarkan sudut kosinus antarvektor. Validasi keakuratan sistem diuji menggunakan 30 sampel data jawaban riil siswa Kelas VIII pada mata pelajaran Ilmu Pengetahuan Alam (IPA). Tingkat presisi model diukur menggunakan metrik *Mean Absolute Error* (MAE) dengan menyandingkan skor otomatis sistem terhadap nilai aktual dari guru. Hasil pengujian menunjukkan sistem berhasil memperoleh nilai *Mean Absolute Error* (MAE) sebesar 3,87 dari skala 0–100. Angka kesalahan yang berada di bawah ambang batas toleransi 5,00 ini membuktikan bahwa algoritma *Cosine Similarity* memiliki tingkat kesalahan yang rendah dan layak diandalkan sebagai instrumen bantu evaluasi akademik di sekolah.

Kata Kunci: *Cosine Similarity*, *Natural Language Processing* (NLP), Penilaian Esai Otomatis, *Computer Based Test* (CBT), *Mean Absolute Error* (MAE).

Abstract– The implementation of conventional paper-based essay examinations at SMP Srijaya Negara encounters operational constraints, particularly regarding prolonged grading durations, calculation error risks, and subjective biases. To address these issues, a web-based *Computer-Based Test* (CBT) platform featuring automated essay grading was designed and engineered utilizing *Natural Language Processing* (NLP) and the *Cosine Similarity* algorithm to ensure objective text matching. Developed under the *Rapid Application Development* (RAD) framework to optimize efficiency, this system runs on PHP, the *Laravel* framework, and a *MySQL* database backend. For text manipulation, the NLP module executes four primary *preprocessing* stages: *case folding*, *tokenizing*, *stopword filtering*, and *stemming*. The normalized texts are subsequently mapped into numerical vectors using the *Bag of Words* (BoW) method to determine similarity scores based on the cosine angle between vectors. System accuracy was validated utilizing 30 authentic eighth-grade science essay responses, evaluated via *Mean Absolute Error* (MAE) statistics against manual teacher grades. The test results indicate that the system successfully achieved a *Mean Absolute Error* (MAE) value of 3.87 on a scale of 0–100. This error rate, which falls below the tolerance threshold of 5.00, demonstrates that the *Cosine Similarity* algorithm has a low error rate and is reliable as a supporting instrument for academic evaluation in schools.

Keywords: *Cosine Similarity*, *Natural Language Processing* (NLP), *Automated Essay Grading*, *Computer Based Test* (CBT), *Mean Absolute Error* (MAE).

1. PENDAHULUAN

Pesatnya inovasi di bidang teknologi informasi saat ini mendorong banyak kemajuan, namun di sisi lain masih terdapat sejumlah proses yang belum sepenuhnya terotomatisasi dan masih dilakukan secara manual. Dalam dunia pendidikan, penerapan teknologi ini telah berkembang pada tahap otomatisasi yang disertai dengan bantuan *Artificial Intelligence* (AI) untuk mendukung kegiatan proses pembelajaran [1]. Salah satu bentuk perwujudan transformasi ini adalah pemanfaatan *Computer Based Test* (CBT). Sebagai metode evaluasi digital, *Computer Based Test* (CBT) memanfaatkan perangkat komputer sebagai sarana inti untuk melangsungkan ujian sekaligus memproses hasil nilainya [2]. Implementasi sistem CBT di lingkungan sekolah membawa dampak positif yang sangat signifikan terhadap efisiensi manajemen evaluasi. Jika dibandingkan dengan metode konvensional, penerapan sistem CBT memungkinkan seluruh rangkaian pelaksanaan ujian, mulai dari distribusi soal hingga

pengumpulan berkas jawaban, dapat berjalan secara jauh lebih cepat, presisi, dan terorganisir [3]. Tidak hanya berfokus pada kecepatan teknis saat pengerjaan ujian, sistem CBT juga mempermudah proses pengolahan data nilai siswa secara otomatis, simultan, dan real-time. Keunggulan lain yang tidak kalah penting terletak pada pemanfaatan platform *Computer Based Test* (CBT) mampu meminimalisasi celah kecurangan serta kebocoran materi ujian yang kerap melanda evaluasi berbasis kertas. Selain itu, sistem digital ini efektif mengeliminasi kendala keterbatasan kuantitas soal, mencegah risiko kerusakan fisik berkas, dan mampu menyajikan skor evaluasi secara *real-time* pasca pengerjaan [4].

Walaupun digitalisasi sistem ujian telah menunjukkan berbagai keunggulan dalam meningkatkan efisiensi dan efektivitas proses evaluasi pembelajaran, implementasinya di lapangan masih belum sepenuhnya merata di berbagai satuan pendidikan [3]. Banyak sekolah menengah hingga saat ini masih menggunakan metode konvensional berbasis kertas sebagai sarana utama dalam pelaksanaan ujian. Ketergantungan terhadap media cetak tersebut menimbulkan sejumlah kendala operasional, seperti lamanya waktu yang dibutuhkan untuk proses pemeriksaan lembar jawaban, tingginya kebutuhan biaya untuk pengadaan dan distribusi kertas dalam jumlah besar, serta meningkatnya potensi kesalahan manusia dalam proses penilaian, misalnya kesalahan perhitungan skor atau risiko hilangnya dokumen ujian. Kondisi ini menjadi semakin kompleks ketika bentuk evaluasi yang digunakan berupa soal esai, karena guru dituntut untuk membaca, memahami, serta menganalisis jawaban siswa secara mendalam satu per satu, bukan sekadar mencocokkan jawaban objektif. Akibatnya, proses penilaian menjadi lebih lambat dan kurang efisien, terutama pada sekolah dengan jumlah peserta didik yang besar.

Situasi tersebut juga terjadi di SMP Srijaya Negara yang hingga saat ini masih menerapkan sistem ujian berbasis kertas dalam kegiatan evaluasi akademik. Dalam praktiknya, seluruh koreksi terhadap lembar jawaban esai sejauh ini masih dikerjakan secara manual oleh guru pengajar, sehingga membutuhkan waktu yang cukup lama dalam penyusunan nilai akhir. Hal ini berdampak pada keterlambatan penyampaian hasil evaluasi kepada siswa. Selain itu, metode penilaian manual berpotensi menimbulkan variasi subjektivitas antar guru. Di sisi lain, sekolah juga mengalami kesulitan dalam mendeteksi tingkat kesamaan jawaban antar siswa atau indikasi kecurangan, terutama ketika jawaban disusun dengan redaksi berbeda namun memiliki makna yang serupa. Kondisi ini membuat proses evaluasi menjadi kurang konsisten dan tidak selalu dapat dilakukan secara cepat dan objektif.

Untuk mengatasi kesenjangan operasional dan standarisasi tersebut, diperlukan sebuah terobosan berupa sistem ujian pintar yang mampu melakukan penilaian esai secara otomatis, objektif, dan konsisten. Salah satu pendekatan ilmiah yang sangat relevan untuk menyelesaikan permasalahan penafsiran teks ini adalah pemanfaatan sub-bidang dari teknologi kecerdasan artifisial. *Natural Language Processing* (NLP) merupakan salah satu ilmu kecerdasan buatan (*Artificial Intelligence*) yang berfokus penuh pada pengembangan kemampuan sistem komputer dalam memahami, mengurai, serta menganalisis bahasa alami manusia secara logis [5]. Dalam konteks evaluasi pendidikan, teknologi *Natural Language Processing* dapat diintegrasikan ke dalam sistem ujian untuk mengukur derajat kemiripan semantik maupun leksikal antara dokumen teks jawaban yang diinput oleh siswa dengan dokumen kunci jawaban yang telah ditetapkan oleh guru. Melalui pendekatan *Artificial Intelligence* dalam pemrosesan bahasa alami ini, sistem mampu memangkas beban kerja administratif guru secara signifikan, menghilangkan unsur bias personal, sekaligus meningkatkan validitas serta transparansi hasil evaluasi akademik secara menyeluruh [6].

Di dalam *Natural Language Processing*, salah satu metode komputasi yang terbukti sangat tangguh dan efisien sebagai instrumen untuk menghitung persentase kesamaan dokumen teks adalah algoritma *Cosine Similarity*. Algoritma *Cosine Similarity* merupakan sebuah metode yang berfungsi untuk mengukur tingkat kesamaan (*similarity*) antara dua objek. Secara teknis, perhitungan dalam metode ini didasarkan pada pendekatan *vector space similarity method*, di mana dua buah dokumen atau objek direpresentasikan ke dalam bentuk vektor [7]. Dalam skenario penilaian esai otomatis, algoritma *Cosine Similarity* memetakan frekuensi kemunculan kata dari teks jawaban siswa menjadi sebuah vektor, kemudian membandingkannya dengan vektor dari teks kunci jawaban resmi milik guru. Melalui perhitungan matematis perkalian titik (*dot product*) yang presisi terhadap panjang vektor tersebut, sistem dapat langsung memberikan skor penilaian secara otomatis dan seketika berdasarkan persentase tingkat kemiripan tekstual yang dihasilkan [8]. Implementasi algoritma *Cosine Similarity* ini diharapkan mampu menciptakan standarisasi penilaian ujian esai yang tidak hanya berjalan jauh lebih cepat, tetapi juga adil, akurat, dan konsisten dari awal hingga akhir pemeriksaan.

Berdasarkan seluruh pemaparan latar belakang masalah, urgensi perbaikan sistem evaluasi, serta potensi besar dari solusi teknologi kecerdasan buatan yang telah diuraikan, fokus utama dari penelitian ini adalah mengonstruksi platform ujian baru dengan skema *Computer Based Test* (CBT) yang modern, yang dilengkapi dengan kemampuan analisis penilaian soal esai otomatis menggunakan pendekatan *Natural Language Processing* (NLP) serta keandalan algoritma *Cosine Similarity*. Melalui integrasi sistem ini, diharapkan SMP Srijaya Negara

dapat mengoptimalkan tata kelola ujian sekolah dan membebaskan tenaga pendidik dari hambatan koreksi manual yang tidak efisien.

2. METODE PENELITIAN

2.1 Metode Pengumpulan Data

Tahap akuisisi dataset pada riset ini dirancang guna memperoleh parameter data yang akurat dan komprehensif, baik berdasarkan sisi kondisi riil di lapangan maupun landasan teori yang mendukung sistem. Pendekatan dalam pengumpulan data pada penelitian ini dikelompokkan ke dalam dua kategori utama, meliputi:

a. Wawancara (*Interview*)

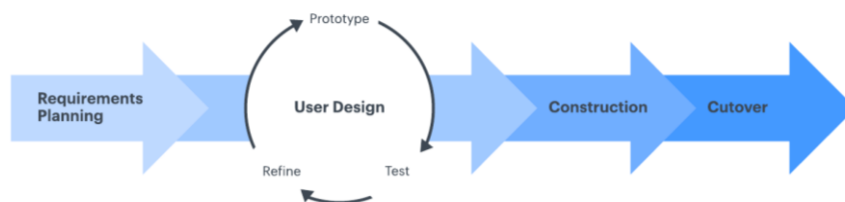
Proses wawancara dilakukan secara langsung dan tatap muka bersama kepala sekolah serta guru mata pelajaran di SMP Srijaya Negara. Kegiatan ini bertujuan untuk memetakan alur pelaksanaan ujian yang saat ini sedang berjalan, khususnya pada pelaksanaan ujian esai yang masih mengandalkan media kertas (*Paper Based Test*) serta proses penilaiannya yang masih bersifat manual. Melalui diskusi mendalam ini, diperoleh informasi mengenai kendala utama yang dihadapi guru seperti lamanya waktu koreksi dan adanya potensi subjektivitas pemberian skor serta poin-poin kebutuhan sistem baru yang diharapkan dapat membantu proses penilaian agar berjalan secara lebih cepat, objektif, dan efektif.

b. Kajian Literatur (*Literature Review*)

Kajian literatur dilaksanakan dengan cara mengumpulkan, menelaah, dan mempelajari berbagai macam sumber referensi akademik terpercaya seperti jurnal ilmiah, buku teks, draf skripsi, serta artikel ilmiah dari penelitian terdahulu yang relevan. Fokus utama pada tahapan ini adalah memperkuat landasan teori dan pemahaman teknis mengenai konsep *Computer Based Test* (CBT) sebagai platform pengujian digital, teknik *Natural Language Processing* (NLP) untuk metode pra-pemrosesan teks, metode *Bag of Words* (BoW) sebagai metode pemetaan karakteristik *term* ke dalam representasi spasial numerik, serta implementasi algoritma *Cosine Similarity* yang digunakan sebagai mesin utama untuk menghitung derajat kemiripan pada penilaian esai otomatis.

2.2 Kerangka Kerja *Rapid Application Development* (RAD)

Pendekatan *Rapid Application Development* (RAD) adalah sebuah model dalam perancangan dan pengembangan perangkat lunak yang memiliki tahapan terstruktur untuk membangun sistem. Metode ini berfokus pada kecepatan, ketepatan, dan akurasi dalam proses pengembangan agar dapat memenuhi kebutuhan pengguna atau pemilik sistem [9]. Metode ini juga bersifat inkremental sehingga sesuai digunakan pada proyek yang harus diselesaikan dalam waktu relatif singkat. Tujuan utama dari RAD adalah menghasilkan sistem yang sesuai dengan kebutuhan serta harapan pengguna [10]. Secara operasional, metodologi RAD ini terbagi menjadi empat tahapan utama [11]:



Gambar 1. Kerangka Kerja *Rapid Application Development* (RAD)

a. Rencana Kebutuhan (*Requirements Planning*)

Identifikasi dan pembedahan masalah pada model ujian tradisional dilakukan melalui diskusi interaktif antara penulis dengan pihak sekolah selaku pemakai sistem pada fase ini, sekaligus menentukan spesifikasi kebutuhan yang diperlukan dalam pengembangan sistem CBT penilaian esai otomatis. Karena tahap tersebut bertindak sebagai fondasi awal penentu keberhasilan sistem, maka komunikasi antara penulis dan pengguna harus berlangsung secara komunikatif, jelas, serta efektif.

b. Desain Pengguna (*User Design*)

Tahap ini merupakan proses penyusunan dan pemodelan rancangan arsitektur sistem yang diusulkan. Prosesnya mencakup perancangan basis data, alur kerja sistem, hingga pembuatan prototipe antarmuka pengguna agar sesuai dengan kebutuhan, dapat diimplementasikan dengan mudah, dan mampu menyelesaikan permasalahan pengoreksian di sekolah.

c. Konstruksi (*Construction*)

Tahap ini adalah proses pembangunan serta pengodean sistem berdasarkan dokumentasi rancangan yang telah dibuat pada fase sebelumnya. Proses dimulai dengan penulisan baris kode program untuk mengubah desain mentah menjadi sebuah aplikasi utuh yang siap digunakan. Dalam menyusun baris kode aplikasi, bahasa pemrograman PHP dan *framework* Laravel dipilih oleh peneliti sebagai fondasi utamanya.

d. Implementasi Sistem (Cutover)

Pada tahapan final ini, sistem CBT yang telah selesai diproduksi wajib melewati proses uji coba secara menyeluruh. Seluruh elemen fungsional, integrasi data, dan akurasi mesin koreksi otomatis diperiksa secara mendalam untuk memantau stabilitas kinerja sistem sekaligus memastikan kesesuaian hasil *output*nya.

2.3 Alur Pemrosesan Teks (Natural Language Processing)

Teknologi *Natural Language Processing* (NLP) diartikan sebagai pendekatan komputasi rekayasa digital guna mendalami sekaligus membedah karakteristik bahasa alami. *Natural Language Processing* (NLP) berfokus pada interaksi antara komputer dan bahasa manusia, dengan tujuan agar mesin mampu memproses, memahami, dan menafsirkan teks atau ucapan secara otomatis untuk menghasilkan informasi yang bermanfaat [12]. Pemrosesan dokumen teks berbahasa Indonesia melalui *Natural Language Processing* (NLP) memerlukan tahapan *text preprocessing* antara lain:

a. Case Folding

Tahapan ini berfokus pada penyamaan seluruh karakter huruf dalam teks agar dikonversi secara seragam menjadi bentuk huruf kecil (*lowercase*).

b. Tokenizing

Proses memotong rangkaian kalimat menjadi satuan kata tunggal sekaligus mengeliminasi tanda baca, angka, dan karakter pembatas.

c. Filtering (Stopword Removal)

Tahap ini berfungsi untuk mengeliminasi kosakata pelengkap atau stopwords yang berfrekuensi tinggi namun tidak memberikan kontribusi signifikan terhadap substansi semantik dokumen.

d. Stemming

Tahapan ini mengonversi bentuk kata berimbuhan ke istilah kata dasar (*root word*) melalui eliminasi afiks awal maupun akhir pada kata tersebut.

2.4 Bag of Word (BoW)

Guna mempermudah pemrosesan komputasional, metode *Bag of Words* (BoW) diterapkan sebagai langkah praktis untuk mengonversi data teks menjadi deretan vektor angka. Pada dasarnya, metode ini bekerja dengan menghitung seberapa sering suatu kata muncul dalam seluruh dokumen yang dianalisis [13]. *Bag of Words* (BoW) digunakan untuk mengubah teks menjadi angka dengan menghitung berapa kali setiap kata muncul. Pada metode tersebut, dokumen teks ditransformasikan ke bentuk larik numerik, yang mana tiap nilai menunjukkan jumlah kemunculan kata tertentu tanpa memperhatikan urutan atau makna kata tersebut. Hasilnya biasanya berupa matriks yang memiliki banyak nilai nol, dan digunakan sebagai dasar untuk menghitung kemiripan teks, seperti pada algoritma cosine similarity [14].

2.5 Alur Algoritma Cosine Similarity dan Evaluasi Mean Absolute Error (MAE)

Setelah dokumen teks melalui fase *Natural Language Processing* (NLP) dan menyisakan kumpulan kata dasar bersih, fitur teks tersebut direpresentasikan ke dalam bentuk vektor numerik menerapkan multidimensi model *Bag of Words* (BoW) berdasarkan frekuensi kemunculan setiap kata unik. Algoritma ini bekerja dengan menerapkan konsep ruang vektor (*Vector Space Model*), di mana setiap dokumen akan diubah menjadi bentuk vektor berdasarkan kumpulan kata yang terkandung di dalamnya. Dalam ranah *text mining*, Cosine Similarity kerap diandalkan untuk menganalisis sekaligus membandingkan teks, sehingga sistem mampu mengidentifikasi hubungan atau kedekatan makna antara satu teks dengan teks lainnya [15]. Formulasi matematis perhitungan Cosine Similarity didefinisikan pada Persamaan (1). Rumus yang digunakan dalam perhitungan Cosine Similarity dapat dinyatakan sebagai berikut [6]:

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

Keterangan :

A= Vektor A

B= Vektor B

A_i= Nilai bobot term ke-i pada vektor A

B_i = Nilai Bobot term ke- i pada vektor B

i = Indeks term pada kalimat

n = Akumulasi jumlah kata keseluruhan

Untuk mengevaluasi keandalan, performa, dan validitas algoritma *Cosine Similarity* tersebut dalam meniru pola penilaian riil guru manusia, dilakukan pengujian kuantitatif terhadap 30 sampel data jawaban esai siswa kelas VIII di SMP Srijaya Negara. Diterapkan sebagai sebuah instrumen untuk mengalkulasi tingkat presisi dari suatu model estimasi maupun prediksi. Metrik *Mean Absolute Error* (MAE) mengekspresikan indeks rata-rata eror mutlak yang membandingkan performa estimasi aplikasi terhadap angka *ground truth* (skor akademik dari staf pengajar atau pakar) [16]. Formulasi matematika perhitungan MAE dijabarkan pada Persamaan (2):

$$MAE = \frac{1}{n} \sum_{i=1}^n |f_i - y_i| \quad (2)$$

Keterangan:

f_i = Nilai prediksi (*Predicted value*)

y_i = Nilai aktual (*Actual value*)

n = Jumlah data (*Number of data points*)

3. HASIL DAN PEMBAHASAN

Bab berikut menjabarkan realisasi rancangan perangkat lunak serta pembahasan mendalam mengenai performa sistem *Computer Based Test* (CBT) yang mengintegrasikan fungsionalitas penilaian esai otomatis di SMP Srijaya Negara. Rangkaian pembahasan diuraikan secara terstruktur, dimulai dari analisis transisi sistem evaluasi konvensional menuju arsitektur digital usulan, perwujudan seluruh komponen antarmuka aplikasi yang dibangun menggunakan framework Laravel, hingga tahapan pengujian fungsionalitas sistem. Lebih lanjut, analisis eksperimen kuantitatif dijabarkan untuk membedah akurasi komputasi algoritma *Cosine Similarity* berbasis *Natural Language Processing* (NLP) melalui representasi model *Bag of Words* (BoW). Seluruh hasil penilaian otomatis dari mesin komputasi kemudian dievaluasi secara langsung terhadap skor aktual dari guru mata pelajaran menggunakan metrik statistik *Mean Absolute Error* (MAE) guna membuktikan validitas, tingkat presisi, serta kelayakan platform ujian digital ini dalam menyamai karakteristik pola penilaian manusia secara objektif.

3.1 Alur yang Diusulkan

Sistem berjalan menunjukkan bahwa evaluasi ujian esai saat ini masih mengandalkan media kertas secara konvensional. Guru harus memeriksa lembar jawaban satu per satu secara manual dengan membandingkannya dengan kunci jawaban, sehingga memakan waktu lama serta rentan terhadap subjektivitas skor akibat variasi susunan kalimat siswa.

Sebagai solusinya, sistem yang diusulkan adalah platform *Computer Based Test* (CBT) berbasis web dinamis dengan tiga hak akses utama. Administrator mengelola data master seperti akun, kelas, dan mata pelajaran; Guru bertugas mengelola paket soal beserta teks kunci jawaban serta meninjau hasil rekap nilai; sedangkan Siswa mengakses sistem secara online untuk mengerjakan soal, di mana jawaban esai yang dikirimkan akan langsung dikoreksi secara otomatis secara *real-time* oleh mesin aplikasi.

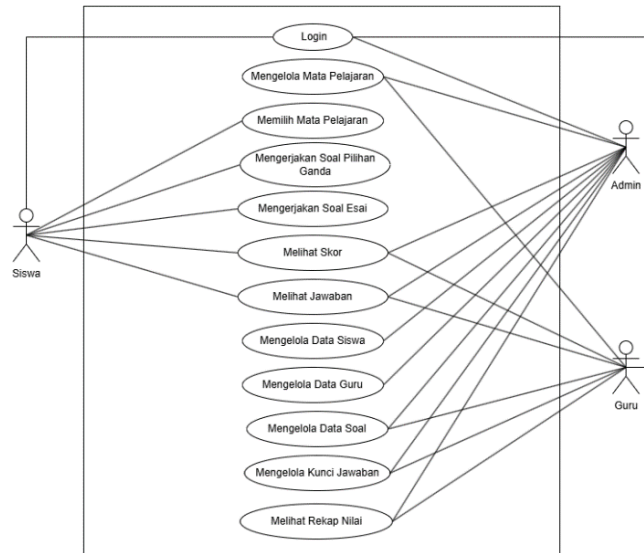
3.2 Perancangan Sistem

Fase ini berfokus pada penyusunan rancangan arsitektur *Computer Based Test* (CBT) di SMP Srijaya Negara guna menyajikan peta visual yang komprehensif terkait struktur dan alur kerja aplikasi yang hendak dibangun. Melalui skema ini, pihak tenaga pendidik beserta administrator sekolah dapat memperoleh pemahaman yang mendalam mengenai fungsionalitas sistem sebelum melangkah ke tahap implementasi. Demi menjamin blueprint yang sistematis dan relevan dengan kebutuhan, pemodelan dilakukan dengan memanfaatkan *Unified Modeling Language* (UML). Sebagai sebuah metodologi perancangan berorientasi objek, UML bertindak sebagai instrumen krusial dalam menstandarisasi desain perangkat lunak. Adapun visualisasi arsitektur sistem dalam penelitian ini dipetakan melalui beberapa model, yang mengintegrasikan skema *use case diagram*, *activity diagram*, serta *class diagram* [17].

3.2.1 Use Case Diagram

Visualisasi dan pemetaan komponen dari sistem rancangan dapat diwujudkan melalui pemanfaatan bermacam-macam tipe diagram yang tersedia di dalam *Unified Modeling Language* (UML), salah satunya adalah *Use Case*

Diagram. Use Case Diagram ini menggambarkan interaksi antara satu atau lebih aktor dengan sistem, serta membantu dalam memahami fungsi-fungsi yang terdapat dalam sistem informasi yang sedang dikembangkan [18]. Dalam *Use Case*, aktor merupakan entitas yang berperan dalam berinteraksi untuk menjalankan tugas atau fungsi tertentu pada sistem penilaian esai ini. Berdasarkan hasil analisis kebutuhan, aplikasi *Computer Based Test* (CBT) di SMP Srijaya Negara melibatkan tiga aktor utama, yaitu Admin, Guru, dan Siswa, dengan *Use Case Diagram* berikut:



Gambar 2. *Use Case Diagram*

Alur interaksi operasional beserta ragam fitur fungsional yang menjembatani tiga pengguna utama (Admin, Guru, dan Siswa) di dalam sistem CBT diilustrasikan secara terstruktur oleh *use case diagram* ini. Hak pemakaian sistem baru akan diberikan secara penuh setelah ketiga kategori aktor tersebut menuntaskan proses otentikasi login sebagai bentuk validasi akses pengguna.

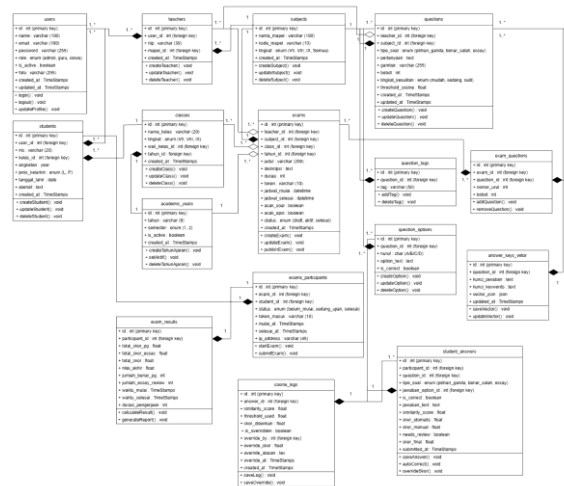
Siswa sebagai pengguna utama dalam pelaksanaan ujian memiliki akses untuk memilih mata pelajaran yang akan dikerjakan. Dalam sistem ini, siswa dapat mengerjakan soal pilihan ganda maupun soal esai (yang nantinya akan diproses menggunakan algoritma *Cosine Similarity*). Setelah ujian selesai, siswa dapat melihat skor yang diperoleh serta pembahasan setiap soal sebagai bahan evaluasi pembelajaran.

Guru memiliki peran penting dalam pengelolaan aspek akademik ujian. Tugas guru meliputi pengelolaan data soal serta pengelolaan kunci jawaban yang menjadi acuan dalam penilaian sistem. Selain itu, guru juga berwenang untuk melihat rekap nilai guna memantau perkembangan nilai seluruh siswa secara keseluruhan. Beberapa fitur seperti melihat skor dan pembahasan juga dapat diakses oleh guru untuk melakukan validasi terhadap hasil ujian siswa.

Admin bertanggung jawab penuh terhadap pengelolaan data master agar sistem dapat berjalan dengan optimal. Tugas admin meliputi pengelolaan mata pelajaran, data siswa, dan data guru. Dengan demikian, admin memastikan seluruh pengguna dan komponen mata pelajaran telah terdaftar dengan benar dalam basis data sistem CBT.

3.2.2 Class Diagram

Class Diagram merupakan salah satu diagram struktur statis yang digunakan untuk menggambarkan struktur serta hubungan antar kelas. Diagram ini dimanfaatkan untuk merepresentasikan objek-objek dalam dunia nyata ke dalam bentuk sistem yang akan dikembangkan [19]. *Class Diagram* juga menunjukkan bagaimana setiap elemen dalam sistem CBT saling berinteraksi, termasuk relasi antar kelas pengguna (Users), pengelolaan ujian (Exams), serta berbagai jenis hubungan lainnya yang membentuk keseluruhan arsitektur perangkat lunak.



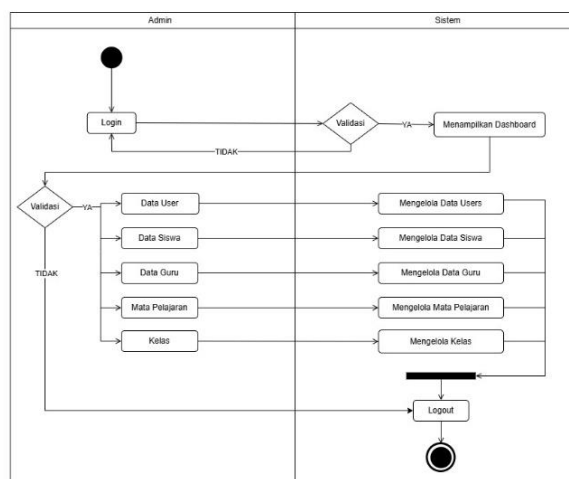
Gambar 3. Class Diagram

Class Diagram tersebut merepresentasikan arsitektur data pada sistem *Computer Based Test* yang mencakup entitas utama seperti *Users*, *Teachers*, *Students*, *Classes*, *Subjects*, dan *Exams*. Guru berperan dalam mengelola instrumen ujian yang meliputi *Questions* (bank soal) dan *Question_Options*, sedangkan setiap siswa terhubung dengan kelas dan tahun ajaran tertentu untuk mendukung pengorganisasian data akademik.

Fokus utama dari diagram ini adalah mekanisme penilaian otomatis, di mana jawaban siswa yang tersimpan pada *Student_Answers* diproses menggunakan algoritma *Cosine Similarity*. Entitas *Answer_Keys_Vector* digunakan untuk menyimpan representasi kunci jawaban yang diperlukan dalam proses *Natural Language Processing* (NLP), sementara hasil perhitungannya dicatat pada *Cosine_Logs* yang memuat skor otomatis. Seluruh hasil penilaian kemudian direkap dalam *Exam_Results* sebagai laporan nilai ujian yang transparan. Secara keseluruhan, struktur ini dirancang untuk mendukung pengelolaan data ujian serta proses penilaian esai secara sistematis dan terorganisir.

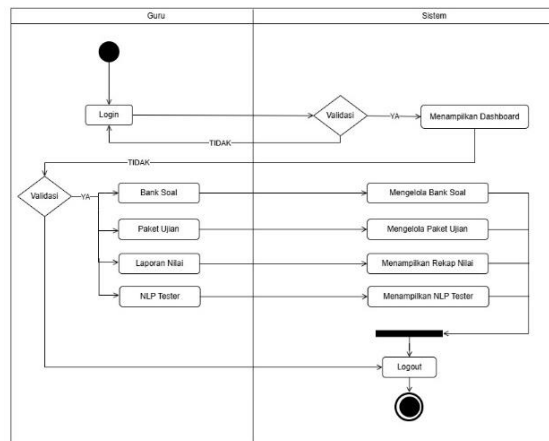
3.2.3 Activity Diagram

Sebagai instrumen pemodelan, *activity diagram* berfungsi mengabstraksikan skema rangkaian tindakan dan tahapan kegiatan yang terjadi di dalam sistem perangkat lunak [20].



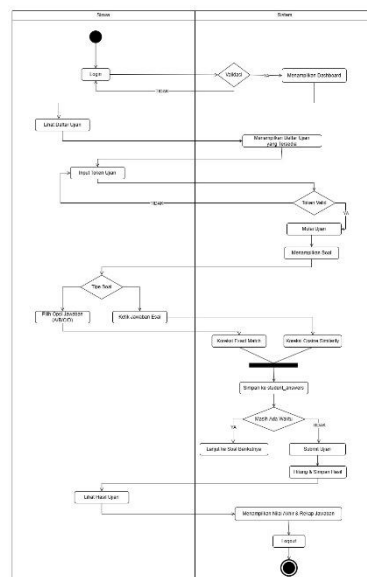
Gambar 4. Diagram Aktivitas Admin

Activity Diagram ini menggambarkan alur proses pengelolaan data oleh Admin dalam sistem *Computer Based Test*. Proses dimulai ketika Admin melakukan *Login*, yang kemudian divalidasi oleh sistem untuk menampilkan Halaman Dashboard. Jika validasi berhasil, Admin dapat mengakses berbagai fitur pengelolaan data utama seperti *Data User*, *Siswa*, *Guru*, *Mata Pelajaran*, hingga *Kelas*. Sistem akan memproses setiap input data sesuai menu yang dipilih oleh Admin. Alur ini diakhiri dengan proses *Logout* yang menutup sesi akses pengguna.



Gambar 5. Diagram Aktivitas Guru

Activity Diagram ini memetakan alur kerja Guru dalam mengelola ujian hingga melakukan validasi penilaian esai. Proses dimulai dari Login yang divalidasi oleh sistem untuk menampilkan *Dashboard*. Guru memiliki akses untuk mengelola Bank Soal, Paket Ujian, Laporan Nilai dan NLP Tester.



Gambar 6. Diagram Aktivitas Siswa

Activity Diagram ini menyajikan alur pengerjaan ujian oleh Siswa di dalam sistem *Computer Based Test*. Proses dimulai dengan *Login* dan validasi token untuk mengakses daftar ujian yang tersedia. Setelah memasukkan Token Ujian yang valid, siswa dapat mulai mengerjakan soal yang ditampilkan oleh sistem.

Fitur utama dalam diagram ini koreksi *Cosine Similarity*. Seluruh jawaban disimpan secara otomatis ke dalam tabel *student_answers*. Setelah ujian dikirim (*submit*), sistem akan menghitung dan menyimpan hasil akhir sehingga siswa dapat langsung Lihat Hasil Ujian yang menampilkan nilai akhir dan rekap jawaban. Alur ditutup dengan proses *Logout* untuk mengakhiri sesi ujian.

3.3 Implementasi Antarmuka Sistem

Memasuki fase konstruksi dalam kerangka kerja *Rapid Application Development* (RAD), platform *Computer Based Test* (CBT) sukses direalisasikan ke dalam bentuk aplikasi berbasis web ditopang oleh *framework* Laravel. Desain antarmuka sistem mengadopsi konsep *responsive web design* guna memastikan fleksibilitas tampilan yang adaptif di berbagai dimensi layar, baik ketika dioperasikan via komputer ataupun perangkat seluler. Di samping itu, aspek kemudahan pemakaian (*usability*) menjadi prioritas dalam perancangan *interface* ini demi menjamin kelancaran navigasi bagi administrator, tenaga pendidik, maupun siswa saat mengoperasikan sistem.

Implementasi antarmuka sistem terdiri atas beberapa halaman utama yang mendukung seluruh proses pelaksanaan ujian.



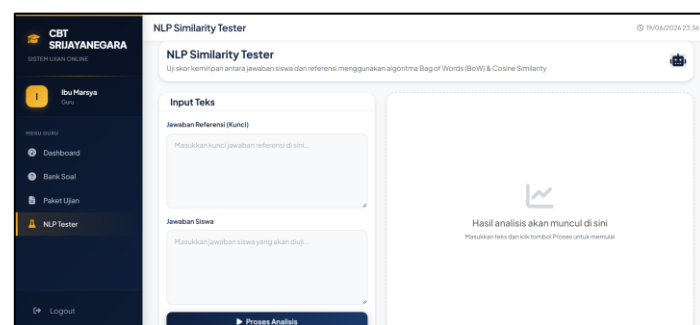
Gambar 7. Antarmuka Halaman Utama

Halaman pertama yang ditampilkan adalah halaman *dashboard*, sebagaimana ditunjukkan pada Gambar 7. Halaman ini berfungsi sebagai pusat informasi dan navigasi utama sistem. Selain itu, halaman dasbor juga menampilkan informasi ringkas yang disajikan dalam bentuk kartu statistik sehingga pengguna dapat memperoleh gambaran kondisi sistem secara cepat ketika pertama kali masuk ke aplikasi.



Gambar 8. Antarmuka Pengerjaan Ujian Esai

Halaman berikutnya merupakan halaman pengerjaan ujian esai yang digunakan oleh siswa saat mengikuti ujian, sebagaimana diilustrasikan pada Gambar 8, antarmuka halaman ini menyediakan kolom input teks tempat siswa mengisikan jawaban esai mereka secara langsung sesuai instruksi soal. Tata letak visual halaman dirancang ke dalam dua panel utama: panel pertama memuat teks soal beserta ruang pengerjaan, sementara panel kedua menyediakan kontrol navigasi angka yang mempermudah siswa beralih dari satu pertanyaan ke pertanyaan lainnya. Selain itu, sistem juga menyediakan penghitung waktu mundur (*countdown timer*) yang berfungsi membantu siswa memantau sisa waktu ujian. Setelah seluruh jawaban selesai dikerjakan, siswa dapat mengirimkan jawaban melalui tombol *submit* sehingga data jawaban akan tersimpan dan diproses oleh sistem.



Gambar 9. Antarmuka NLP Tester

Halaman berikutnya adalah halaman NLP Tester yang diperuntukkan bagi guru sebagai sarana pengujian proses penilaian otomatis, sebagaimana ditunjukkan pada Gambar 9. Melalui halaman ini, guru dapat memasukkan kunci jawaban dan contoh jawaban siswa untuk melihat hasil perhitungan kemiripan teks yang dilakukan oleh sistem. Selain menampilkan nilai kemiripan, halaman ini juga memperlihatkan hasil dari setiap

tahapan pemrosesan teks, mulai dari *case folding*, *tokenizing*, *filtering*, hingga *stemming*. Informasi tersebut membantu guru memahami proses yang dilakukan sistem sebelum menghasilkan nilai akhir.

3.4 Hasil Penerapan NLP dan Representasi Vektor

Dalam menguji keandalan performanya, platform CBT ini memanfaatkan data riil yang berasal dari hasil penilaian akademis mata pelajaran Ilmu Pengetahuan Alam (IPA) Kelas VIII di SMP Srijaya Negara. Pengujian komparatif ini mencakup total 150 instrumen data jawaban teks dari 30 siswa untuk 5 butir soal berbentuk esai. Hasil transformasi tekstual secara terstruktur untuk satu sampel data jawaban siswa mengenai materi pesawat sederhana disajikan pada Tabel 1.

Tabel 1. Hasil Pengolahan *Text Preprocessing* pada Sampel Jawaban Esai

Tahap NLP	Teks Jawaban Siswa	Teks Kunci Jawaban Guru
Kalimat Awal	Pesawat sederhana yang berupa papan/bidang yang dibuat miring. Tujuannya adalah untuk memperkecil usaha saat memindahkan beban yang berat.	Bidang miring adalah pesawat sederhana yang berupa papan/bidang yang dibuat miring. Tujuannya adalah untuk memperkecil usaha saat memindahkan beban yang berat.
<i>Case Folding</i>	pesawat sederhana yang berupa papan/bidang yang dibuat miring. tujuannya adalah untuk memperkecil usaha saat memindahkan beban yang berat.	bidang miring adalah pesawat sederhana yang berupa papan/bidang yang dibuat miring. tujuannya adalah untuk memperkecil usaha saat memindahkan beban yang berat.
<i>Tokenizing</i>	[pesawat], [sederhana], [yang], [berupa], [papan], [bidang], [yang], [dibuat], [miring], [tujuannya], [adalah], [untuk], [memperkecil], [usaha], [saat], [memindahkan], [beban], [yang], [berat]	[bidang], [miring], [adalah], [pesawat], [sederhana], [yang], [berupa], [papan], [bidang], [yang], [dibuat], [miring], [tujuannya], [adalah], [untuk], [memperkecil], [usaha], [saat], [memindahkan], [beban], [yang], [berat]
<i>Filtering</i>	[pesawat], [sederhana], [papan], [bidang], [dibuat], [miring], [tujuannya], [memperkecil], [usaha], [memindahkan], [beban], [berat]	[bidang], [miring], [pesawat], [sederhana], [papan], [bidang], [dibuat], [miring], [tujuannya], [memperkecil], [usaha], [memindahkan], [beban], [berat]
<i>Stemming</i>	[pesawat], [sederhana], [papan], [bidang], [buat], [miring], [tuju], [kecil], [usaha], [pindah], [beban], [berat]	[bidang], [miring], [pesawat], [sederhana], [papan], [bidang], [buat], [miring], [tuju], [kecil], [usaha], [pindah], [beban], [berat]

Berdasarkan visualisasi data pada Tabel 1, sistem terbukti mampu melakukan reduksi kata berimbuhan menjadi kata dasar secara presisi, seperti perubahan kata "memperkecil" menjadi "kecil" serta eliminasi kata-kata fungsional (stopwords) tanpa membuang kata kunci esensial. Kumpulan kata dasar bersih tersebut kemudian dipetakan ke dalam model representasi *Bag of Words* (BoW) untuk dihitung frekuensi kemunculannya. Struktur pemetaan bobot kata unik ke dalam bentuk vektor numerik spasial dijabarkan pada Tabel 2.

Tabel 2. Matriks Frekuensi Kata Baku pada Model *Bag of Words*

No	Kosakata Unik (<i>Term</i>)	Bobot Vektor Jawaban Siswa (A)	Bobot Vektor Kunci Jawaban (B)
1	bidang	1	2
2	miring	1	2
3	pesawat	1	1
4	sederhana	1	1
5	papan	1	1
6	buat	1	1
7	tuju	1	1
8	kecil	1	1
9	usaha	1	1
10	pindah	1	1
11	beban	1	1
12	berat	1	1

Melalui representasi spasial pada Tabel 2, terbentuk sepasang nilai array biner berupa Vektor $A = [1,1,1,1,1,1,1,1,1,1,1,1]$ dan Vektor $B = [2,2,1,1,1,1,1,1,1,1,1,1]$. Tahapan perhitungan kecocokan nilai *Cosine Similarity* diselesaikan melalui operasi matematika berikut:

a. Menghitung Dot Product ($A \cdot B$)

$$A \cdot B = (1 \times 2) + (1 \times 2) + (1 \times 1) \times 10$$

$$A \cdot B = 14$$

b. Menghitung Panjang Vektor A ($\|A\|$)

$$\|A\| = \sqrt{1^2 \times 12}$$

$$\|A\| = \sqrt{12} = 3,464$$

c. Menghitung Panjang Vektor B ($\|B\|$)

$$\|B\| = \sqrt{(2^2 + 2^2 + (1^2 \times 10))}$$

$$\|B\| = \sqrt{18} = 4,243$$

d. Nilai Skor Kemiripan (*Similarity Score*) dihitung menggunakan Persamaan (1):

$$\text{Similarity}(A, B) = \frac{14}{3,464 \times 4,243} = \frac{14}{14,697} = 0,953$$

Hasil kosinus sebesar 0,953 pada perhitungan di atas mengindikasikan tingkat akurasi kemiripan semantik yang sangat tinggi (rentang $0,7 < SIM \leq 1$). Sistem menilai jawaban ini berkategori "Benar" karena siswa berhasil memuat seluruh inti makna, meskipun melewati susunan kata frasa "bidang miring adalah" di awal kalimat jawaban.

3.5 Pengujian Mean Absolute Error (MAE)

Pengujian akurasi dilakukan untuk mengetahui performa dan tingkat kesalahan (error) dari algoritma *Cosine Similarity* berbasis *Natural Language Processing* (NLP) yang diterapkan pada sistem dalam menilai jawaban esai siswa. Sampel data yang digunakan dalam pengujian ini terdiri dari 150 instrumen data jawaban teks yang diperoleh dari hasil pengerjaan 5 butir soal esai oleh 30 siswa pada evaluasi akademis mata pelajaran Ilmu Pengetahuan Alam untuk jenjang Kelas VIII pada SMP Srijaya Negara. Setiap jawaban siswa dikonversi menjadi vektor *Bag of Words* (BoW) kemudian dihitung kemiripannya dengan vektor kunci jawaban untuk menghasilkan nilai otomatis dari sistem. Berikut merupakan data perbandingan penilaian manual yang telah dilakukan guru dengan penilaian otomatis dari sistem pada Tabel 3.

Tabel 3. Data Perbandingan Penilaian Guru dan Penilaian Sistem

No	Siswa	Nilai Sistem (f_i)	Nilai Aktual / Guru (y_i)	Selisih Mutlak
1	S1	90,6	95	4,4
2	S2	84,8	90	5,2
3	S3	83,0	80	3,0
4	S4	91,8	95	3,2
5	S5	80,9	75	5,9
6	S6	90,0	90	0,0
7	S7	90,1	95	4,9
8	S8	87,4	90	2,6
9	S9	93,3	90	3,3
10	S10	89,1	85	4,1
11	S11	94,0	90	4,0
12	S12	91,7	95	3,3
13	S13	90,1	90	0,1
14	S14	94,6	100	5,4
15	S15	87,5	90	2,5
16	S16	75,1	75	0,1
17	S17	86,3	75	11,3
18	S18	92,0	85	7,0
19	S19	81,2	70	11,2
20	S20	77,0	75	2,0
21	S21	71,0	75	4,0
22	S22	77,7	70	7,7
23	S23	89,5	90	0,5

24	S24	88,8	85	3,8
25	S25	74,0	75	1,0
26	S26	70,7	70	0,7
27	S27	82,0	80	2,0
28	S28	73,8	70	3,8
29	S29	81,8	75	6,8
30	S30	82,6	85	2,4
Total				116,2

Dari hasil komparasi langsung antara nilai otomatis sistem f_i dengan skor aktual yang diberikan oleh guru y_i terhadap 30 sampel data tersebut, diperoleh akumulasi total selisih mutlak ($\sum |f_i - y_i|$) sebesar 116,2. Nilai rata-rata kesalahan absolut sistem kemudian dihitung dengan mengimplementasikan rumus *Mean Absolute Error* (MAE) pada Persamaan (2) sebagai berikut:

$$MAE = \frac{1}{n} \sum_{i=1}^n |f_i - y_i|$$

$$MAE = \frac{116,2}{30} = 3,87$$

Melalui perhitungan matematis di atas, didapatkan nilai kesalahan mutlak rata-rata (MAE) sebesar 3,87. Berdasarkan kriteria standarisasi pengujian statistik, perolehan nilai MAE yang berada di bawah angka 5,00 mengindikasikan bahwa sistem penilaian esai otomatis menggunakan algoritma *Cosine Similarity* berbasis model BoW memiliki performa akurasi yang Sangat Tinggi. Hasil deviasi yang sangat minim ini (rata-rata hanya berselisih $\pm 3,87$ poin pada skala 0–100) membuktikan bahwa kedekatan hasil penilaian sistem sudah sangat presisi dalam menyamai pola penilaian subjektif manusia (guru), sehingga platform CBT ini valid dan sangat layak diimplementasikan.

4. KESIMPULAN

Setelah melewati fase perancangan hingga serangkaian evaluasi berbasis data kuantitatif pada aplikasi ujian *Computer Based Test* (CBT) dengan fitur penilai esai otomatis di lingkungan SMP Srijaya Negara, dapat ditarik beberapa kesimpulan utama yang secara langsung menjawab seluruh rumusan permasalahan di dalam penelitian ini. Pertama, penelitian tugas akhir ini telah berhasil membangun sebuah platform aplikasi CBT digital berbasis web, proses rekayasa perangkat lunak ini bersandar pada kerangka kerja *Rapid Application Development* (RAD), yang mana implementasi teknisnya memanfaatkan kombinasi *software framework* Laravel, bahasa PHP, dan MySQL. Pengembangan perangkat lunak ini sukses mentransformasikan paradigma pelaksanaan evaluasi konvensional bermedia kertas (*Paper Based Test*) yang hingga kini masih diterapkan di sekolah menjadi sebuah ekosistem digital yang adaptif, aman, dan terstruktur. Melalui pembagian hak akses yang terintegrasi secara modular antara administrator, guru, dan siswa, platform usulan ini mampu memangkas alur operasional pelaksanaan ujian secara signifikan, mulai dari tahap distribusi bank soal oleh guru hingga tahap rekapitulasi data nilai siswa yang kini dapat berjalan secara otomatis dan real-time.

Penerapan teknologi *Natural Language Processing* (NLP) yang diintegrasikan melibatkan empat alur inti pra-pemrosesan data teks (*text preprocessing*), yang mencakup tahapan *case folding*, *tokenizing*, *filtering* (*stopword removal*), hingga *stemming*. terbukti mampu menyisakan komponen kata-kata dasar informatif (*root words*) yang valid tanpa membuang substansi esensi makna dari jawaban tersebut. Melalui pemetaan frekuensi kata menggunakan model *Bag of Words* (BoW) ke dalam *Vector Space Model*, algoritma *Cosine Similarity* berhasil menyelesaikan kendala utama pendidik dalam mengenali variasi parafrase atau susunan kalimat siswa yang berbeda namun sebenarnya mengandung inti makna yang serupa dengan kunci jawaban guru. Evaluasi komputasi spasial terhadap sudut kosinus menunjukkan bahwa algoritma ini mampu mendeteksi kedekatan leksikal secara objektif dan adil, sehingga mampu mengeliminasi bias personal dan subjektivitas penilaian yang sering kali muncul pada metode koreksi manual akibat faktor kelelahan fisik maupun penurunan fokus guru di sekolah.

Guna mengukur tingkat presisi dari sistem koreksi esai otomatis yang telah dibangun, pengujian dilakukan dengan memanfaatkan 30 sampel lembar jawaban Ilmu Pengetahuan Alam untuk jenjang Kelas VIII pada SMP Srijaya Negara. Prosedur evaluasi ini dijalankan dengan mengomparasikan skor keluaran platform terhadap nilai konvensional yang diberikan langsung oleh guru pengampu selaku tolok ukur utama. Melalui

implementasi metrik *Mean Absolute Error* (MAE), hasil kalkulasi menunjukkan tingkat kesalahan rata-rata yang diperoleh adalah sebesar 3,87.

Nilai *Mean Absolute Error* (MAE) tersebut menunjukkan bahwa perbedaan antara hasil penilaian sistem dan penilaian guru relatif kecil. Semakin rendah nilai *Mean Absolute Error* (MAE) yang diperoleh, semakin kecil selisih yang terbentuk antara hasil penilaian otomatis aplikasi dengan skor aktual. Melalui perolehan nilai *Mean Absolute Error* (MAE) sebesar 3,87 yang berada di bawah standar kriteria pengujian statistik sebesar 5,00, sistem terbukti mampu menghasilkan penilaian yang akurat serta konsisten dalam mengevaluasi jawaban esai siswa. Hasil ini menunjukkan bahwa metode *Natural Language Processing* (NLP) dan algoritma *Cosine Similarity* yang diterapkan dapat digunakan sebagai pendukung proses penilaian esai secara otomatis. Oleh karena itu, sistem yang dikembangkan dinilai layak untuk membantu guru dalam mempercepat proses koreksi sekaligus meningkatkan efisiensi pelaksanaan evaluasi pembelajaran.

UCAPAN TERIMAKASIH

Peneliti menyampaikan penghargaan tertinggi kepada pihak SMP Srijaya Negara atas kerja sama dan bantuannya, baik dalam hal perizinan studi maupun pemenuhan kebutuhan data jawaban esai siswa. Tak lupa pula, terima kasih juga disampaikan untuk Politeknik Negeri Sriwijaya beserta dosen pembimbing yang telah memberikan dukungan penuh guna menyelesaikan tugas akhir ini dengan lancar.

DAFTAR PUSTAKA

- [1] M. Maulani, "RANCANG BANGUN APLIKASI COMPUTER BASED TEST BERBASIS WEB PADA SMPN 1 KATAPANG KABUPATEN BANDUNG SELATAN," 2021.
- [2] D. Setiawan, "Penerapan Exampro Sebagai Pendukung CBT (Computer Base Test)," Nov. 2021.
- [3] D. Wijayanto, "Manajemen Evaluasi Pembelajaran Computer Based Testing (CBT) di SMP Kapanewon Wonosari Gunungkidul," Oct. 2022, [Online]. Available: <http://jurnal.ustjogja.ac.id/index.php/mmp>
- [4] I. P. Y. Agus Ariwanta, I. G. T. Eka Saputra, N. P. E. Apriyanthi, I. M. A. O. Gunawan, and G. Indrawan, "Analisis Kepuasan Pengguna Menggunakan Metode EUCS Pada Sistem Computer Based Test di Institusi Pendidikan," *Journal of Computer System and Informatics (JoSYC)*, vol. 4, no. 4, pp. 942–951, Aug. 2023, doi: 10.47065/josyc.v4i4.3752.
- [5] F. Rumaisa, Y. Puspitarani, A. Rosita, A. Zakiah, and S. Violina, "Penerapan Natural Language Processing (NLP) Di Bidang Pendidikan," Dec. 2021.
- [6] R. R. Rafi, R. A. Ramadhani, and A. Sanjaya, "PEMANFAATAN ALGORITMA COSINE SIMILARITY UNTUK MENGKOREKSI UJIAN ESAI," *JUPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 10, no. 3, pp. 2242–2254, Aug. 2025, doi: 10.29100/jupi.v10i3.8058.
- [7] F. Ri. Kurniawan and F. M. Arif, "IMPLEMENTASI TEXT MINING MENGGUNAKAN METODE COSINE SIMILARITY UNTUK KLASIFIKASI KONTEN BERITA DI POSTINGAN GRUP FACEBOOK INFO LANTAS DAN KRIMINAL PASURUAN," *JAMI: Jurnal Ahli Muda Indonesia*, vol. 3, no. 1, pp. 9–17, Jun. 2022, doi: 10.46510/jami.v3i1.41.
- [8] N. Rokhman, P. A. Maulan, and N. A. Wirahuda, "Analisis Penilaian Esai Secara Otomatis Menggunakan Natural Language Processing (NLP) dan Cosine Similarity," *Go Infotech: Jurnal Ilmiah STMIK AUB*, vol. 31, no. 1, pp. 41–52, Jun. 2025, doi: 10.36309/goi.v31i1.359.
- [9] Y. Natadisastro, N. Safaat Harahap, and T. Darmizal, "MALCOM: Indonesian Journal of Machine Learning and Computer Science Implementation of Large Language Models in the Design of Android-Based Computer-Assisted Test Mobile Applications Implementasi Large Language Models dalam Rancang Bangun Aplikasi Computer-Assisted Test Mobile Berbasis Android," vol. 6, no. 1, pp. 52–61, Jan. 2026, doi: 10.57152/malcom.v6i1.2402.
- [10] A. Basri, V. Kuswanto, and A. Leo, "Rancang Bangun Bridging Sistem Pendaftaran dan Aplikasi Mobile Jaminan Kesehatan Nasional (JKN)," 2022.

- [11] H. Rianto and O. Pahlevi, "Pengembangan Billing Online Sistem dengan Menggunakan Metode Rapid Application Development," *JURIKOM (Jurnal Riset Komputer)*, vol. 9, no. 6, p. 2021, Dec. 2022, doi: 10.30865/jurikom.v9i6.5213.
- [12] D. O. Sihombing, "Implementasi Natural Language Processing (NLP) dan Algoritma Cosine Similarity dalam Penilaian Ujian Esai Otomatis," *Jurnal Sistem Komputer dan Informatika (JSON)*, vol. 4, no. 2, p. 396, Dec. 2022, doi: 10.30865/json.v4i2.5374.
- [13] A. A. Firdaus, A. Id Hadiana, and A. K. Ningsih, "Klasifikasi Sentimen pada Aplikasi Shopee Menggunakan Fitur Bag of Word dan Algoritma Random Forest," *R2J*, vol. 6, no. 5, 2024, doi: 10.38035/rj.v6i5.
- [14] R. P. Permata, R. N. Suharto, L. Candra, and A. Julianty, "Towards an Automated Essay Evaluation System NLP Based Text Embeddings and Similarity Metrics," *Jurnal Teknologi Informasi & Komunikasi*, vol. 16, Apr. 2025, doi: 10.31849/digitalzone.v16i1.
- [15] A. Arsad, M. Hamid, and M. Santosa, "PENERAPAN TEKS MINING DAN COSINE SIMILARITY UNTUK MENENTUKAN KESAMAAN DOKUMEN SKRIPSI APPLICATION OF TEXT MINING AND COSINE SIMILARITY TO DETERMINE THE SIMILARITY OF THESIS DOCUMENTS," Apr. 2024.
- [16] A. M. Fawwazrin, A. Arbansyah, and Amin Rezaeipah, "ID Card Printing Time Prediction With Mean Absolute Error in Linear Regression," *JSE Journal of Science and Engineering*, vol. 3, no. 2, pp. 80–84, Jan. 2025, doi: 10.30650/jse.v3i2.3868.
- [17] C. Binangkit, A. Voutama, and N. Heryana, "PEMANFAATAN UML (UNIFIED MODELING LANGUAGE) DALAM PERENCANAAN SISTEM PENGELOLAAN SEWA ALAT MUSIK BERBASIS WEBSITE," 2023.
- [18] R. P. Pangestu and A. Voutama, "PEMANFAATAN UML (UNIFIED MODELLING LANGUAGE) PADA SISTEM PENGELOLAAN ASPIRASI MAHASISWA BERBASIS WEBSITE," 2024.
- [19] P. Putra, V. Pranatawijaya, N. Sari, and Widiarty, "Mobile Android Queue Covid19 Application Registration and Application of Queues at the Mobile," vol. 16, no. 1, 2022, doi: 10.47111/JTI.
- [20] F. Ardhy *et al.*, "Pelatihan Analisis dan Desain Sistem Informasi Menggunakan Unified Modeling Language (UML) di SMK Pelita Madani Kabupaten Pringsewu," *Abdimas Universal*, vol. 5, no. 1, pp. 97–104, Apr. 2023, doi: 10.36277/abdimasuniversal.v5i1.285.