

# Pengembangan Aplikasi Penilaian Otomatis Ujian Online Essay Dengan Menggunakan Metode Cosine Similarity dan Model XLM-RoBERTa (XLM-R) Berbasis Website

Mochamad Faridz Dwi Putra<sup>1</sup>, Ari Kurniawan<sup>2</sup>

Manajemen Informatika, Fakultas Vokasi, Universitas Negeri Surabaya  
Surabaya, Indonesia

[1mochamad.19043@mhs.unesa.ac.id](mailto:mochamad.19043@mhs.unesa.ac.id)

[2arikurniawan@unesa.ac.id](mailto:arikurniawan@unesa.ac.id)

**Abstrak**— Ujian seringkali digunakan sebagai bahan evaluasi hasil akhir untuk mengukur tingkat kemampuan serta pemahaman terhadap peserta didik. Namun dalam pelaksanaannya, ujian masih banyak dilakukan secara manual terutama ujian dengan tipe essay yang membuat peserta ujian harus menyiapkan kertas dan alat tulis sendiri. Sedangkan bagi tenaga pendidik membutuhkan banyak tenaga dan waktu yang lama untuk memeriksa jawaban ujian satu persatu. Dengan menerapkan metode pengembangan Rapid Application Development (RAD) maka penelitian ini dilakukan dengan membuat sistem ujian online dengan tipe ujian essay beserta penilaian secara otomatis. Penilaian dilakukan dengan cara melihat kemiripan jawaban antara jawaban tenaga pendidik dengan jawaban peserta ujian menggunakan metode cosine similarity dan model XLM-RoBERTa. Hasil dari penelitian ini menunjukkan persentase perbandingan penilaian manual dan otomatis menunjukkan bahwa jawaban yang memiliki perhitungan matematika memiliki persentase 60%, sedangkan pada penilaian dengan jawaban tanpa perhitungan matematika hanya berupa text/kalimat serta jawaban yang dapat dibolak balik akan tetapi memiliki arti yang sama memiliki nilai nilai persentase 85%.

**Kata kunci**— Web ujian online, cosine similarity, XLM-R, ujian essay, penilaian otomatis.

**Abstract**— Exams are often used as material for evaluating the final results to measure the level of ability and understanding of students. However, in practice, many exams are still carried out manually, especially exams with the essay type, which require examinees to prepare their own paper and stationery. Meanwhile, educators need a lot of energy and a long time to check exam answers one by one. By applying the Rapid Application Development (RAD) development method, this research was carried out by creating an online exam system with the type of essay exam along with automatic scoring. The assessment was carried out by looking at the similarity of the answers between the answers of the teaching staff and the answers of the examinees using the cosine similarity method and the XLM-RoBERTa model. The results of this study show that the percentage comparison of manual and automatic assessments shows that answers that have mathematical calculations have a percentage of 60%, while the assessment with answers without mathematical calculations is only in the form of text/sentences and answers that can be reversed but have the same meaning have a percentage value of 85%.

**Keywords:** Exam online, cosine similarity, XLM-R, essay exam, automatic ratings.

## I. PENDAHULUAN

Untuk mewujudkan pengembangan sumber daya manusia yang baik dan kooperatif, diperlukan pelatihan yang mendefinisikan cara berpikir manusia, khususnya bagi peserta didik belajar di Lembaga pendidikan tertentu. Seiring dengan berkembangnya zaman diperlukan alat pendidikan yang dapat membantu peserta didik dalam memperoleh pengetahuan serta informasi secara cepat[1].

Dalam mengukur kemampuan peserta didik pada lembaga pendidikan, seringkali tenaga pendidik memberikan ujian kepada peserta didiknya. Namun, dalam praktiknya, ujian kebanyakan dilaksanakan secara manual diatas kertas. Hal ini dapat menyebabkan pemborosan kertas yang berlebihan dan menyebabkan tenaga pendidik menghabiskan banyak waktu untuk memeriksa jawaban hasil ujian[2].

Seiring berjalannya waktu, diperkenalkan ujian yang terkomputerisasi serta dapat memeriksa jawaban secara otomatis dan dapat memunculkan hasil ujian secara langsung tanpa harus diperiksa secara manual oleh tenaga pendidik. Jenis ujian yang umum digunakan saat ini yaitu *multiple choice* dan *essay*. Namun dalam prakteknya saat ini pemeriksaan jawaban secara otomatis masih banyak terdapat pada jenis ujian *multiple choice*, masih belum banyak platform ujian online yang dapat memeriksa jawaban ujian *essay* secara otomatis[3].

Penerapan metode Cosine Similarity dalam penelitian ini digunakan untuk mengukur tingkat kemiripan pada dua jawaban antara jawaban tenaga pendidik dan peserta ujian [4]. Dengan menggabungkan metode XLM-RoBERTa (XLM-R) yang sudah dilatih dengan menggunakan 100 bahasa, termasuk sumber daya rendah seperti Bahasa Indonesia. Hasil dari XLM-R nantinya berupa representasi vektor yang nantinya digunakan dalam perhitungan pada model Cosine Similarity[5].

Berdasarkan penjelasan yang sudah diuraikan diatas penulis membuat sebuah penelitian untuk membuat sistem penilaian otomatis ujian online dengan tipe essay dengan judul “Pengembangan Aplikasi Penilaian Otomatis Ujian Online

Essay Dengan Menggunakan Metode Cosine Similarity Dan Model XLM-RoBERTa (XLM-R) Berbasis Website”. Hasil dari penelitian tersebut berupa aplikasi yang dapat digunakan oleh seluruh tenaga pendidik untuk membuat soal ujian dan memberikan token ujian kepada peserta ujian. Setelah ujian dilaksanakan maka secara otomatis dilakukan pencocokan jawaban antara jawaban yang sudah dikerjakan oleh peserta ujian dengan kunci jawaban yang sudah tersimpan baik bentuk soal ujian *multiple choice* maupun *essay*.

## II. METODE PENELITIAN

### A. Alur Penelitian

Dalam melakukan penelitian ini, terdapat beberapa tahapan perencanaan penelitian dan perancangan sistem yang dapat dilaksanakan. Berikut adalah tahapan-tahapan perencanaan penelitian diatas:

#### 1. Studi Literatur

Yaitu melaksanakan studi literatur untuk menemukan sebuah masalah, mencari dan membaca dari sumber-sumber jurnal ilmiah yang sudah memiliki kredibilitas tinggi serta memiliki keterkaitan dengan topik peneliti.

#### 2. Perancangan Aplikasi

Melakukan perancangan aplikasi nantinya aplikasi dapat berjalan dengan baik dan sesuai dengan alur yang sudah direncanakan.

#### 3. Implementasi

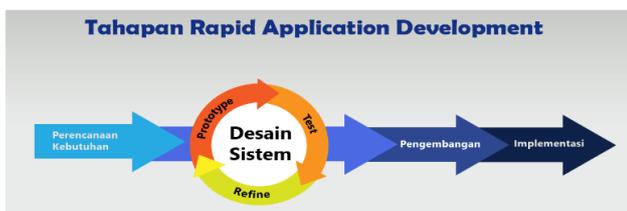
Implementasi dilakukan dengan mengimplementasikan metode Cosine Similarity dan model XLM-RoBERTa (XLM-R) dalam melaksanakan pemeriksaan jawaban ujian jenis *essay*.

#### 4. Testing dan Debugging

Testing dilakukan terhadap aplikasi yang sudah dibuat untuk memastikan aplikasi berjalan sesuai dengan rancangan penelitian yang sudah dibuat.

### B. Tahapan Penelitian

Dalam melakukan penelitian, penulis menggunakan metode *Rapid Application Development* (RAD) yang dibagi menjadi beberapa tahapan seperti pada gambar 1 berikut.



Gambar 1 Metode Rapid Application Development

#### 1. Perancangan Kebutuhan

Yaitu tahapan untuk melakukan identifikasi menentukan tujuan akhir dari pembuatan sebuah produk.

#### 2. Desain Sistem

Yaitu tahapan perbaikan desain secara berulang apabila terdapat ketidaksesuaian terhadap hasil akhir yang akan dibuat.

#### 3. Proses Pengembangan

Yaitu tahapan untuk mengembangkan sistem yang akan dibuat.

#### 4. Implementasi

Yaitu tahapan realisasi dari suatu aplikasi sesuai dengan rancangan, fungsi, visual, dan model desain yang sudah disepakati.

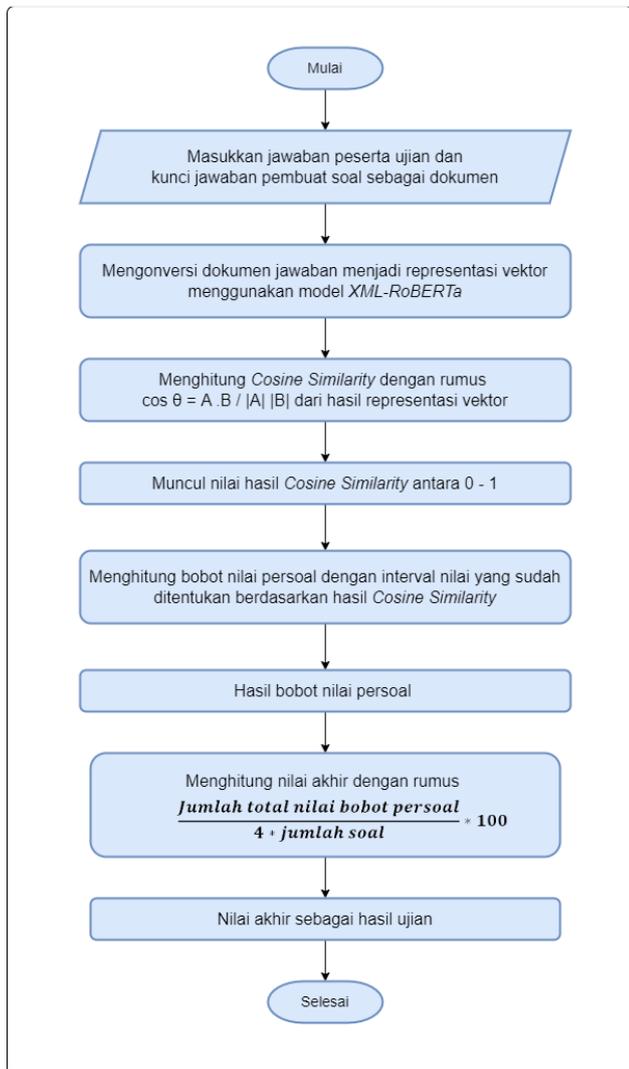
### C. Sumber Data Penelitian

Sumber data yang digunakan pada penelitian ini didapatkan dari kunci jawaban yang sudah dimasukkan pada database oleh tenaga pendidik dan hasil jawaban dari peserta ujian.

### D. Tahapan Penilaian Essay dengan Menggunakan Model XLM-RoBERTa dan Model Cosine Similarity.

Perhitungan otomatis penilaian pada soal essay berdasarkan interval yang sudah ditetapkan oleh penulis. Nantinya hasil representasi vektor dan perhitungan model Cosine Similarity mendapatkan nilai antara 0 hingga 1.

Dari hasil Cosine Similarity akan dilakukan pengecekan sesuai dengan interval yang sudah ditentukan sehingga akan menghasilkan nilai dari tiap soal *essay* tersebut. Seperti yang sudah dijelaskan *flowchart* pada gambar 2 berikut.

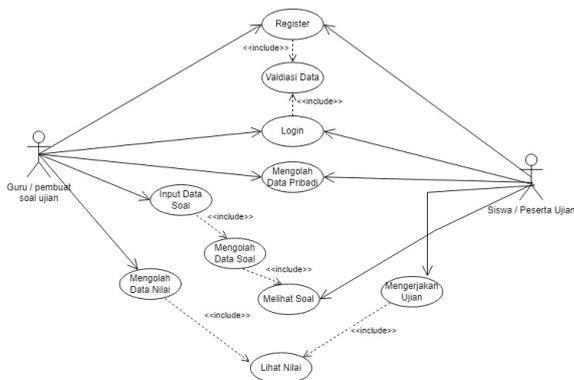


Gambar 2 Flowchart penerapan XLM-RoBERTa dalam penilaian essay

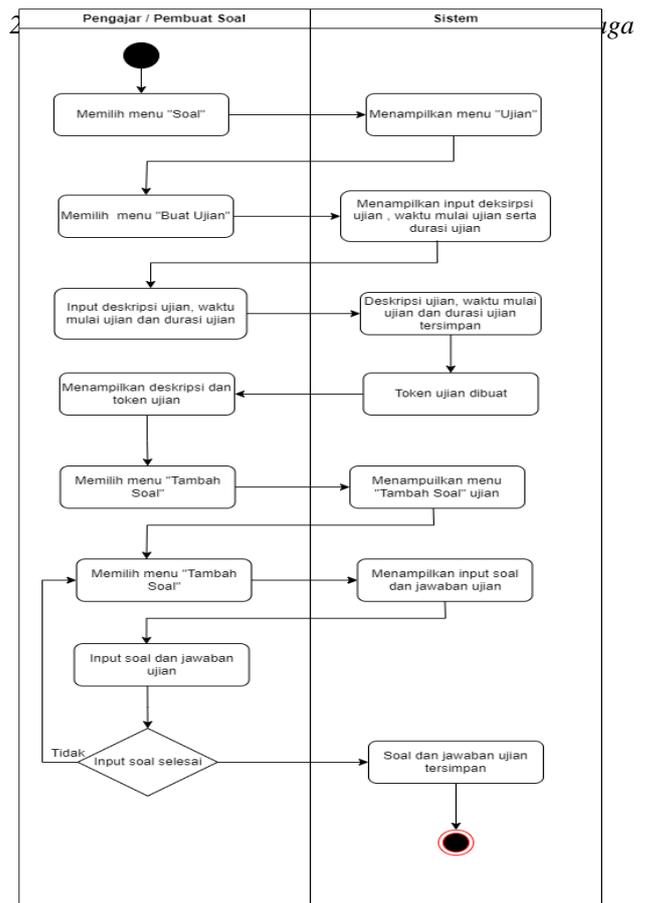
### E. Rancangan Sistem

#### 1. Use Case Diagram

Use Case diagram dibuat untuk memudahkan interaksi antara user pengguna dengan sistem seperti pada gambar a3 berikut.



Gambar 3 Rancangan Use Case diagram



Gambar 4 Activity diagram membuat soal oleh tenaga pendidik

Gambar 4 diatas menjelaskan bagaimana aktivitas membuat soal bagi tenaga pendidik. Baik soal dengan jenis ujian *multiple choice* ataupun *essay*

#### 3. Activity Diagram Mengerjakan Soal Ujian oleh Peserta Ujian

Activity Diagram mengerjakan soal ujian menjelaskan bagaimana aktivitas peserta ujian untuk dapat mengakses aplikasi serta mengikuti ujian dan mendapatkan hasil ujian secara langsung. Seperti yang sudah digambarkan pada gambar 5 berikut.

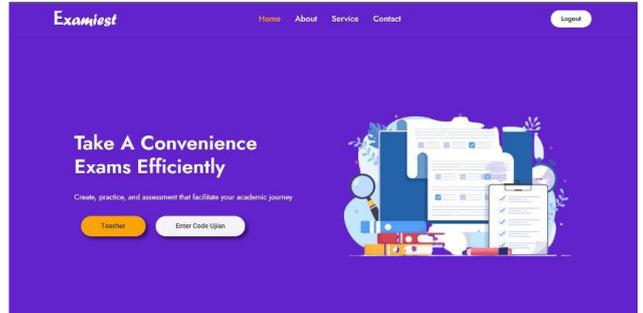
### III. HASIL DAN PEMBAHASAN

Pada bab ini merupakan tahapan menerjemahkan perancangan berdasarkan hasil analisa sebelumnya kedalam bahasa yang dimengerti oleh mesin.

#### A. Hasil User Interface

User interface di dalam sistem ini dibuat dengan menggunakan *HTML5*, *CSS3*, *Javascript*, *JQuery*, *Bootstrap*, dan *framework Laravel*

##### 1. Halaman Landing Page

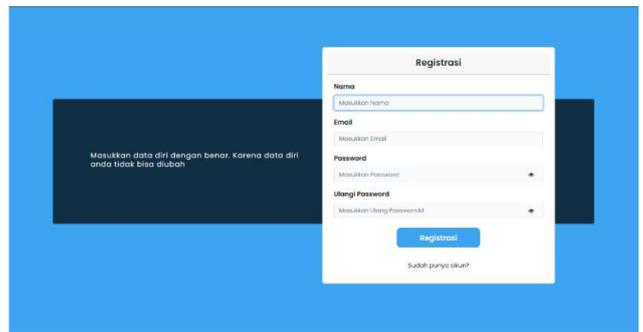


Gambar 8 Halaman landing page

Hasil dari penelitian berupa aplikasi berbasis website dengan nama “Examist” yang memiliki arti dalam bahasa Indonesia yaitu paling teliti. Gambar 8 diatas merupakan tampilan awal dari website Examist.

##### 2. Halaman Registrasi

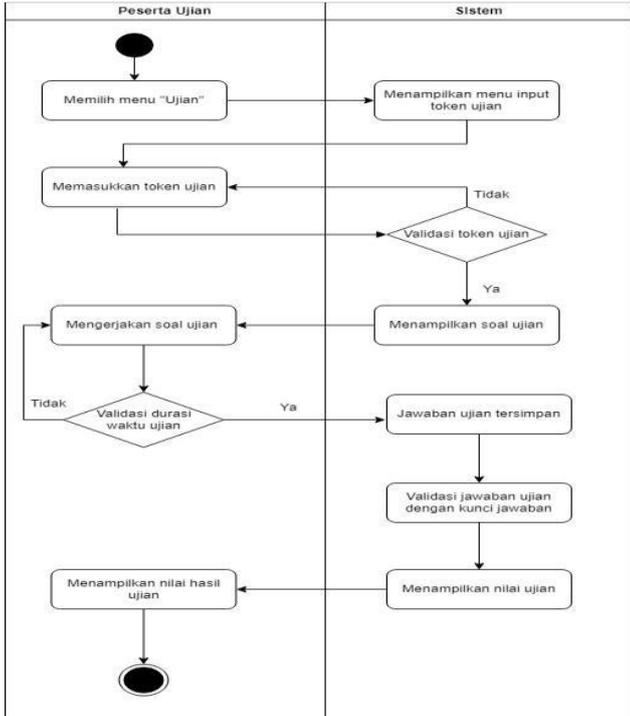
Halaman registrasi digunakan untuk tenaga pendidik dapat membuat akun yang akan digunakan untuk mengkases website untuk membuat dan mengelola hasil ujian.



Gambar 9 Halaman registrasi

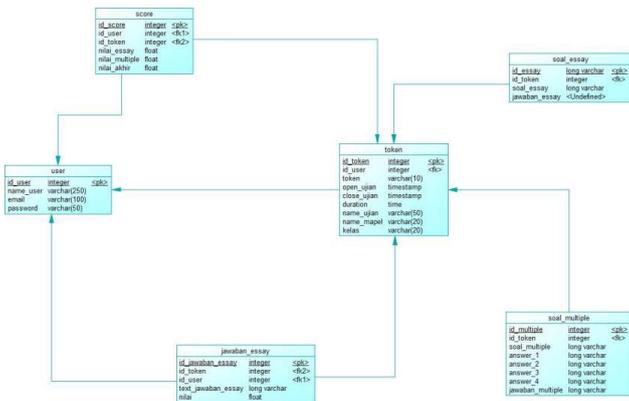
##### 3. Halaman Login

Halaman login digunakan oleh tenaga pendidik apabila sudah mempunyai akses untuk membuat dan mengelola hasil ujian. Halaman login hanya dapat digunakan oleh tenaga pendidik. Untuk peserta ujian tidak memerlukan hak akses dalam mengikuti ujian. Tampilan halaman login sesuai pada gambar 10 berikut.



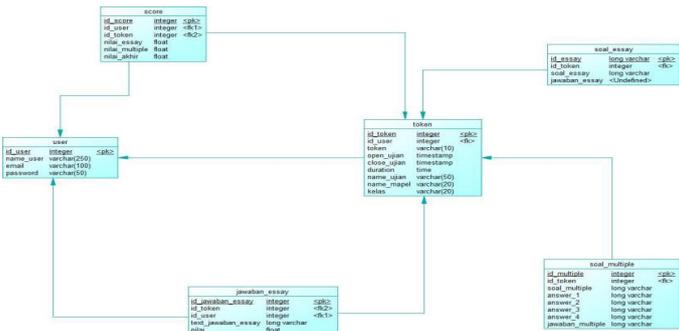
Gambar 5 Activity diagram mengerjakan soal oleh peserta ujian

#### 4. Rancangan Conceptual Data Model (CDM)

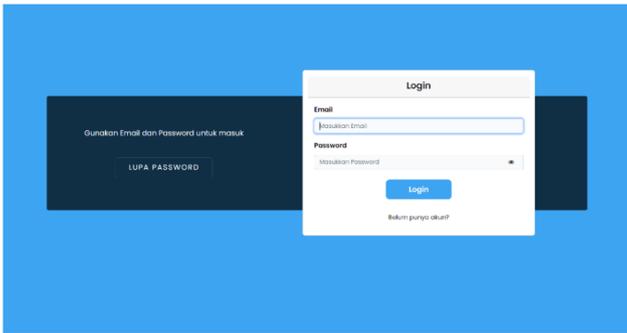


Gambar 6 Rancangan Conceptual Data Model (CDM)

#### 5. Rancangan Physical Data Model (PDM)

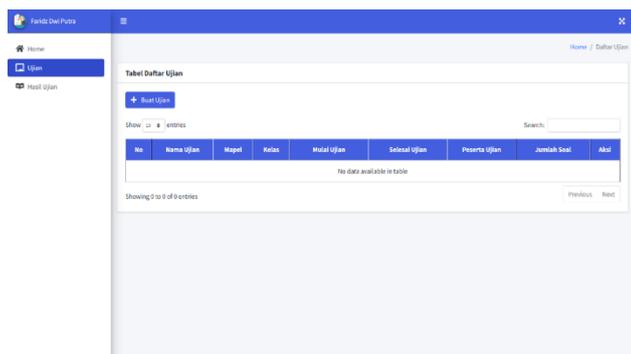


Gambar 7 Physical Data Model (PDM)



Gambar 10 Tampilan halaman login

#### 4. Dashboard Buat Ujian



Gambar 11 Halaman dashboard buat ujian

Gambar 11 diatas menjelaskan tampilan dashboard yang terdapat fitur buat ujian. Tampilan tersebut digunakan untuk membuat ujian sesuai dengan kebutuhan tenaga pendidik.

#### 5. Form Tambah Ujian

Tambah Ujian
✕

---

**Nama Ujian**

**Mata Pelajaran**

**Kelas**

**Mulai Ujian**

**Akhir Ujian**

**Durasi Ujian**

 Menit

\* Duration ujian dalam bentuk menit

**Jumlah soal multiple yang akan tampil**

**Jumlah soal essay yang akan tampil**

**Persentase Nilai Multiple**

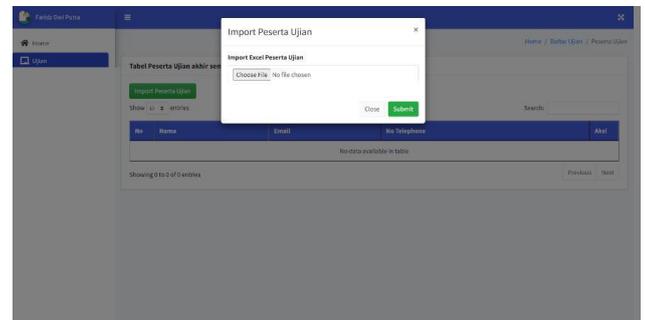
**Persentase Nilai Essay**

Close Simpan

Gambar 12 Form tambah ujian

Gambar 12 adalah tampilan form dari fitur buat ujian yang terdapat di dalam dashboard ujian. Tampilan form buat ujian berisi nama ujian,waktu pelaksanaan ujian serta jumlah soal ujian yang nantinya akan ditampilkan.

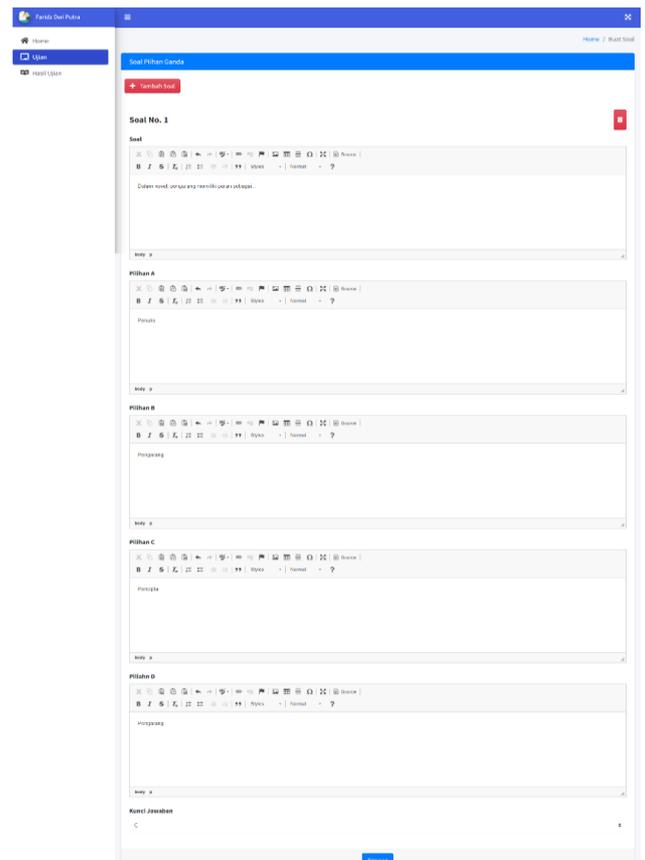
#### 6. Halaman *Import* Peserta Ujian



Gambar 13 Tampilan import peserta ujian

Setelah tenaga pendidik berhasil membuat ujian. Tenaga pendidik harus menambahkan peserta ujian yang mempunyai hak akses terhadap ujian tersebut. Nantinya terdapat token dari setiap peserta setelah ditambahkan kedalam ujian.

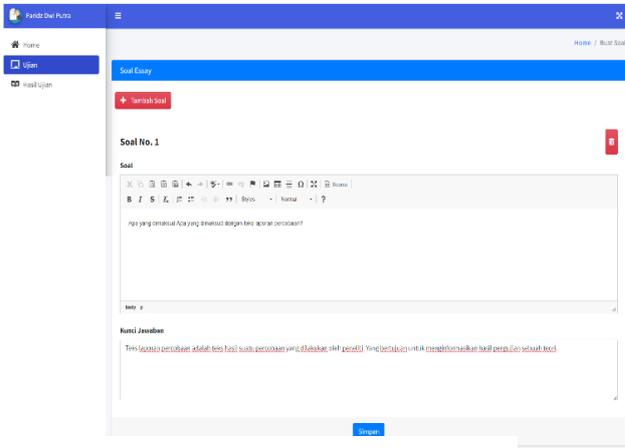
#### 7. Halaman Buat Soal *Multiple Choice*



Gambar 14 Bsuat soal multiple choice

Gambar 14 merupakan tampilan dari tenaga pendidik untuk membuat soal dengan jenis ujian *multiple choice* beserta kunci jawabannya. Nantinya tenaga pendidik dapat menambahkan soal dengan kebutuhan dalam pelaksanaan ujian.

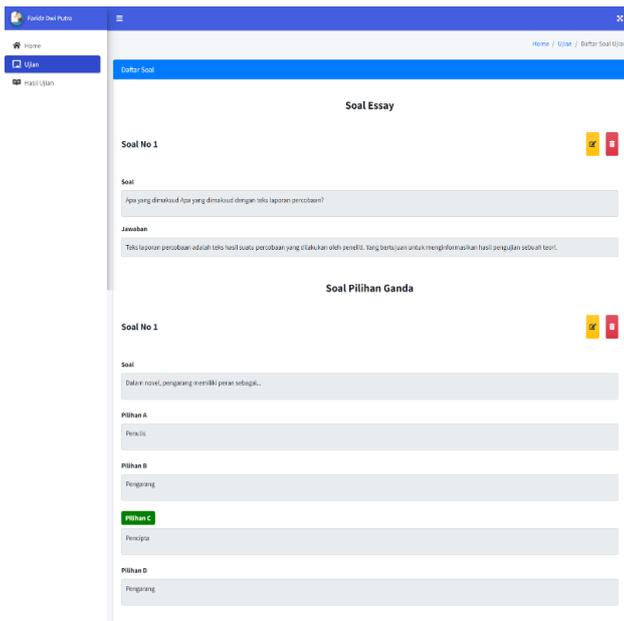
### 8. Halaman Buat Soal Essay



Gambar 15 Buat soal essay

Gambar 15 merupakan tampilan dari halaman untuk membuat soal essay beserta kunci jawaban untuk tenaga pendidik.

### 9. Halaman Detail Soal Ujian



Gambar 16 Detail Hasil Soal Ujian

Setelah tenaga pendidik membuat soal ujian dengan jenis soal *multiple choice* maupun *essay*, tenaga pendidik dapat melihat hasil soal yang sudah dibuat sebelumnya pada menu detail soal ujian. Selain

itu tenaga pendidik dapat mengubah maupun menghapus soal tersebut.

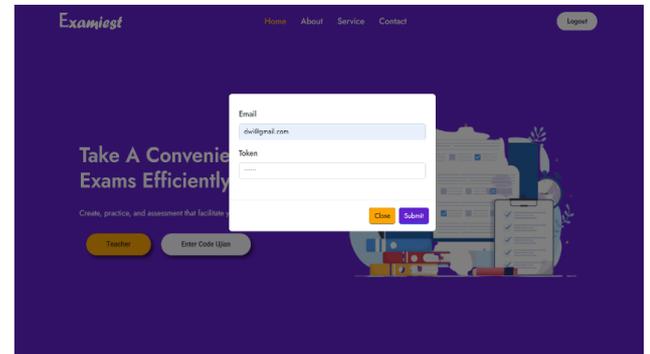
### 10. Email Token Peserta Ujian



Gambar 17 Email token untuk peserta ujian

Peserta dapat mengikuti ujian apabila telah memiliki token yang didapat dari masing-masing email peserta. Nantinya token akan berbeda di setiap email masing-masing peserta ujian.

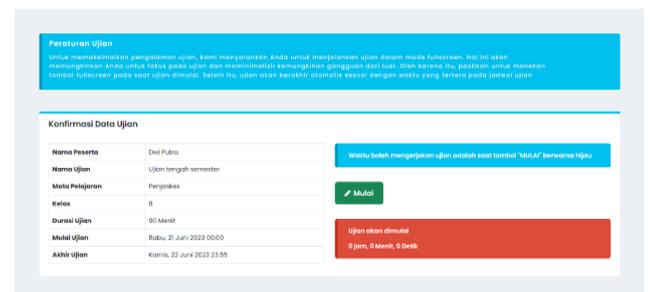
### 11. Form Peserta Ujian



Gambar 18 Form peserta ujian

Gambar 18 diatas merupakan tampilan form ketika peserta ingin mengikuti ujian. Peserta harus mengisi email dan juga token yang sudah didapat dari email masing-masing peserta.

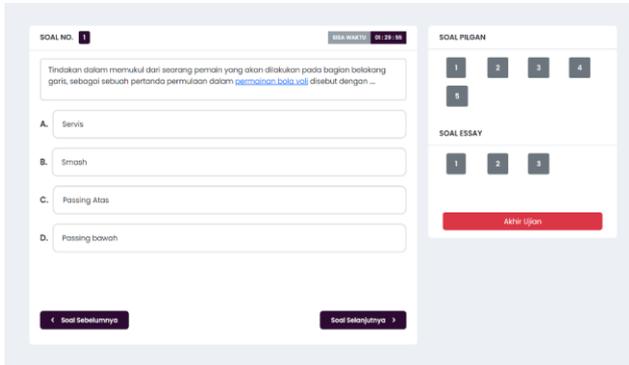
### 12. Halaman Detail Ujian



Gambar 19 Detail informasi ujian

Setelah peserta berhasil mengisi email dan token yang sesuai, maka sistem secara otomatis akan menampilkan detail ujian yang akan diikuti oleh peserta tersebut, yang meliputi nama ujian, waktu ujian dilaksanakan dan durasi ujian sesuai pada gambar 19.

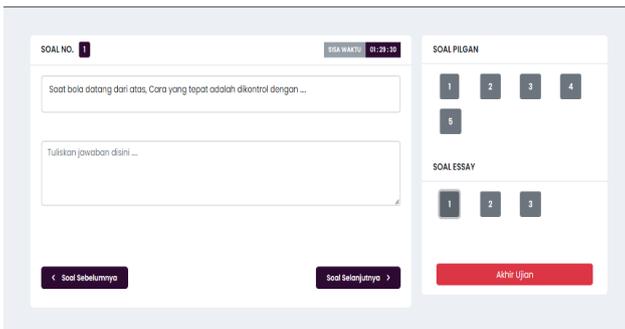
### 13. Tampilan Ujian Jenis Soal *Multiple Choice*



Gambar 20 Tampilan ujian jenis *multiple choice*

Gambar 20 diatas merupakan tampilan dari jenis soal ujian *multiple choice*. Nantinya peserta dapat memilih pilihan jawaban yang sesuai dan dapat berpindah ke nomor soal yang lain.

### 14. Tampilan Ujian Jenis Soal *Essay*



Gambar 21 Tampilan ujian jenis *essay*

Gambar 21 merupakan tampilan ujian dengan jenis soal *essay* yang menampilkan soal beserta kolom uraian jawaban untuk peserta ujian.

### B. Proses Penilaian Ujian

Setelah peserta selesai mengerjakan ujian, maka secara otomatis akan menampilkan hasil ujian secara langsung baik dengan jenis ujian *multiple choice* maupun *essay*.

Untuk penilaian *multiple choice* akan menilai dengan cara menyamakan hasil pilihan jawaban peserta dengan kunci jawaban dari tenaga pendidik. Untuk jenis ujian *essay* penilaian dengan cara mencari kesamaan dari kedua jawaban dengan

menggunakan model XLM-RoBERTa (XLM-R) dan metode Cosine Similarity. Hasil dari cosine similarity nantinya akan dipetakan sesuai interval yang sudah ditentukan peneliti.

Apabila hasil Cosine similarity antara 0.95 hingga 1 maka nilai jawaban soal tersebut adalah 1, apabila hasil similarity bernilai antara 0.89 hingga 0.95 mendapatkan nilai 3, sedangkan apabila hasil similarity bernilai antara 0.75 hingga 0.89 mendapatkan nilai 2 dan apabila hasil kurang dari 0.75 mendapatkan nilai 1.

### C. Pengujian dengan Data Hasil Ujian

Pengujian dilakukan dengan melakukan perbandingan antara penilain manual yang sudah dilakukan oleh guru dari hasil jawaban ujian penilaian sumatif satuan pendidikan tahun 2022/2023 di SMP Islam Terpadu At-Taqwa Surabaya. Lembar jawaban diambil dari 10 peserta kelas IX-E untuk mata pelajaran prakarya.

Pengujian dilakukan pada hasil jawaban yang mengandung perhitungan matematika dan jawaban yang hanya berupa text. Karena pada model XLM-R masih belum bisa melakukan penilaian secara akurat dari hasil perhitungan matematika.

1. Hasil pengujian penilaian *essay* dengan jawaban yang mengandung perhitungan matematika.

| No  | Nama                          | Penilaian |   | Cosine Similarity |
|-----|-------------------------------|-----------|---|-------------------|
|     |                               | M         | O |                   |
| 1.  | Athyya                        | 4         | 4 | 0.95873           |
| 2.  | Diyah Nor Malasari            | 1         | 3 | 0.98560           |
| 3.  | Jasmine                       | 4         | 4 | 0.97992           |
| 4.  | Verghia Balqis Anindya        | 4         | 3 | 0.88805           |
| 5.  | Aura Arsy Diu Cahhyadi        | 4         | 4 | 0.98791           |
| 6.  | Raissa Zandra Widyadhri Putri | 1         | 3 | 0.91517           |
| 7.  | Octavia Tsania                | 4         | 4 | 0.97871           |
| 8.  | Lorelai Talitha Zafir         | 1         | 3 | 0.92699           |
| 9.  | Atania Maghfiratika           | 4         | 4 | 0.98975           |
| 10. | Raihanum Azzahra Zein         | 1         | 4 | 0.98556           |

2. Hasil Pengujian penilaian *essay* dengan jawaban hanya berupa kalimat

| No | Nama | Penilaian |  |
|----|------|-----------|--|
|    |      |           |  |

|     |                               | M | O | Cosine Similarity |
|-----|-------------------------------|---|---|-------------------|
| 1.  | Athyya                        | 2 | 3 | 0.95873           |
| 2.  | Diyah Nor Malasari            | 2 | 3 | 0.98560           |
| 3.  | Jasmine                       | 3 | 4 | 0.97992           |
| 4.  | Verghia Balqis Anindya        | 2 | 3 | 0.88805           |
| 5.  | Aura Arsy Diu Cahyadi         | 3 | 3 | 0.98791           |
| 6.  | Raissa Zandra Widyadhri Putri | 3 | 3 | 0.91517           |
| 7.  | Octavia Tsania                | 4 | 4 | 0.97871           |
| 8.  | Lorelai Talitha Zafir         | 2 | 3 | 0.92699           |
| 9.  | Atania Maghfiratika           | 4 | 4 | 0.98975           |
| 10. | Raihanum Azzahra Zein         | 3 | 4 | 0.98556           |

Dari hasil perbandingan penilain secara manual yang sudah dilakukan oleh tenaga pendidik dengan penilaian otomatis menggunakan XLM-RoBERTa dan cosine similarity diperoleh persentase dari hasil penilaian manual dan otomatis pada soal dengan jawaban yang memiliki perhitungan matematika yaitu memiliki tingkat keakuratan 60%. Dikarenakan penilaian otomatis dengan menggunakan model XLM-RoBERTa dan cosine similarity tidak dapat melakukan perbandingan dokumen yang memiliki perhitungan matematika.

Dari hasil penilaian dengan menggunakan soal yang memiliki jawaban tanpa perhitungan matematika yang hanya berupa text/kalimat serta jawaban yang dapat dibolak balik akan tetapi memiliki arti yang sama. Memiliki tingkat keakuratan 85% karena memiliki kesamaan antara penilaian manual dengan otomatis.

#### IV. KESIMPULAN DAN SARAN

##### A. Kesimpulan

1. Aplikasi ujian online dengan penilaian otomatis pada tipe ujian essay dibuat dengan menggunakan framework Laravel dan library inertia.js. Selain itu model XLM-RoBERTa dan cosine similarity digunakan dalam penilain essay secara otomatis yang dibuat dengan menggunakan bahasa python dengan menggunakan fast api.
2. Hasil aplikasi ujian online dengan penilaian otomatis pada tipe ujian essay dapat digunakan oleh semua

tenaga pendidik dengan cara membuat akun di dalam aplikasi tersebut.

3. Hasil aplikasi penilaian otomatis ujian online essay yang dibangun dapat berjalan sesuai dengan rancangan serta otomatis penilaian ujian essay akan dilakukan setelah peserta ujian selesai melaksanakan ujian. Selain itu aplikasi dapat digunakan oleh semua instansi pendidikan dengan cara tenaga pendidik sudah terdaftar dalam aplikasi tersebut untuk membuat ujian dan menambahkan peserta ujian pada masing-masing ujian.

##### B. Saran

Beberapa saran dan masukan berikut diharapkan dapat memberikan perbaikan sistem pada penelitian selanjutnya, antara lain:

1. Sistem penilaian essay pada penelitian ini yang ada di dalam aplikasi hanya 4 interval saja tidak bisa ditambah ataupun berkurang. Diharapkan pada penelitian selanjutnya diharapkan dapat membuat interval penilaian essay sesuai dengan kebutuhan tenaga pendidik.
2. Pada penelitian ini aplikasi yang sudah dibuat masih belum bisa menangani permasalahan kesalahan penulisan kata(typo). Diharapkan pada penelitian selanjutnya dapat menggabungkan model untuk menangani kesalahan penulisan (typo).
3. Penelitian lebih lanjut diharapkan menggunakan model penilaian essay berdasarkan kesamaan kalimat yang berbeda lalu melakukan perbandingan dengan penelitian ini dapat menemukan mana model yang tepat. Nantinya model tersebut dapat digunakan dalam pemeriksaan kalimat dalam membantu ujian essay.

#### REFERENSI

- [1] Yulianto Erwin. (2020). "Pembangunan Aplikasi Ujian Online Menggunakan Akses Token & Algoritma Simple Random Sampling".
- [2] Kurniawati Endah dan pradnya mega (2020). "Implementasi Algoritma Winnowing Pada Sistem Penilaian Otomatis Jawaban Essay Pada Ujian Online Berbasis Web".
- [3] Aprianto Tio (2017). "Perancangan Sistem Pemeriksaan jawaban Otomatis Ujian Online Multiple Choice Dan Essay Pada Kuliah Mikroprosesor Berbasis Tf-Idf Dan Vektor Space Model Di Stmik Asia Malang".
- [4] Lahitani R A., 2022. "Automated Essay Scoring menggunakan Cosine Similarity pada Penilaian Esai Multi Soal". Vol 22 No 2 (Mei 2022). 107-118.
- [5] Wiciaputra Keenan Yakobus. dkk (2021). "Bilingual Text Classification in English and Indonesian via Transfer Learning using XLMRoBERTa".