

## **PENERAPAN SISTEM PENILAIAN OTOMATIS PADA JAWABAN TES URAIAN MENGGUNAKAN ALGORITMA RABIN KARP UNTUK MATA PELAJARAN INFORMATIKA DI MA KANJENG SEPUH**

**Dimas Tifli Irhami Hamir**

Pendidikan Teknologi Informasi, Fakultas Teknik, Universitas Negeri Surabaya  
Email: [dimas.18001@mhs.unesa.ac.id](mailto:dimas.18001@mhs.unesa.ac.id)

**Ekohariadi**

Pendidikan Teknologi Informasi, Fakultas Teknik, Universitas Negeri Surabaya  
Email: [ekohariadi@unesa.ac.id](mailto:ekohariadi@unesa.ac.id)

### **Abstrak**

Penilaian hasil belajar merupakan faktor yang sangat penting dalam dunia pendidikan, karena terkait dengan pengambilan keputusan atas pencapaian hasil belajar peserta didik setelah mengikuti proses pembelajaran. Penilaian yang berbentuk tes uraian merupakan salah satu teknik yang dapat digunakan dalam proses penilaian hasil belajar. Dalam penerapan penilaian secara otomatis pada jawaban tes uraian, telah dikembangkan berbagai model dan metode yang bervariasi, salah satunya seperti metode *string matching*, dan algoritma yang dapat digunakan dalam melakukan prosesnya adalah algoritma *Rabin Karp* yang merupakan algoritma pencarian *string multi pattern* yang dikembangkan oleh Michael O. Rabin dan Richard M. Karp pada tahun 1987. Media penilaian otomatis pada jawaban tes uraian dalam penelitian ini menggunakan algoritma *Rabin Karp* dengan pendekatan *Synonym Recognition* dan akan dikembangkan dengan metode *Research and Development* (R&D) model Plomp. Telah dilakukan validasi pada media penilaian otomatis oleh para ahli, dan diperoleh hasil persentase nilai validasi sebesar 90,51% yang dapat dinyatakan bahwa media yang telah dikembangkan sangat valid dan layak untuk diterapkan dalam proses penilaian hasil belajar. Hasil perhitungan akhir nilai peserta didik yang dilakukan secara otomatis memperoleh rata-rata sebesar 76,28 dan untuk hasil penilaian secara manual diperoleh rata-rata sebesar 74,15 yang kemudian akan dilakukan uji reliabilitas dan uji kesepakatan antara kedua hasil nilai pada tiap butir soalnya dengan menggunakan koefisien *Cohen's Kappa* dan koefisien *intra-class correlation* (ICC) yang kemudian didapatkan hasil uji dengan kategori *Low Agreement with High Reliability*.

**Kata Kunci:** Evaluasi, Tes Uraian, *Rabin Karp*, *Synonym Recognition*.

### **Abstract**

Assessment of learning outcomes is very important component of education, because it is related to decision making on the achievement of student learning outcomes after participating in the learning process. Assessment in essay form is one technique that can be used in the process of assessing learning outcomes. In the application of automatic assessment of essay test answers, various models and methods have been developed, one of which is a method based on string matching methods, and the algorithm that can be used to do this is the Rabin Karp algorithm which is an algorithm multi-pattern string search developed by Michael O. Rabin and Richard M. Karp in 1987. The automatic assessment tool of essay test answers in this study uses the Rabin Karp algorithm with the *Synonym Recognition*, and will be developed using the *Research and Development* (R&D) method by Plomp. The automatic assessment tool has been validated and obtained a score of 90.51% which can be stated that the media that has been developed is very valid and feasible to be applied in the learning outcome assessment process. The results of the final calculation of student scores which are carried out automatically get an average of 76.28 and for manual assessment results an average of 74.15 is obtained, which will then be tested for reliability and agreement test between the two scores on each item. using *Cohen's Kappa* coefficient and *intra-class correlation coefficient* (ICC) which later found the test results with the category of *Low Agreement with High Reliability*.

**Keywords:** Evaluation, Essay, *Rabin Karp*, *Synonym Recognition*.

### **PENDAHULUAN**

Penilaian hasil belajar merupakan faktor yang sangat penting dalam dunia pendidikan, karena terkait dengan pengambilan keputusan atas pencapaian hasil belajar peserta didik setelah mengikuti proses pembelajaran. Supriyati dan Dudung (2018:1) dalam bukunya menjelaskan bahwa sebelum seorang tenaga pendidik

mengambil keputusan terkait keberhasilan peserta didik dalam mencapai suatu kompetensi maka dibutuhkan data berupa informasi terkait hasil belajar peserta didik, oleh karenanya penilaian hasil belajar merupakan salah satu pilar dalam penerapan kurikulum berbasis kompetensi.

Terdapat banyak teknik yang dapat digunakan sebagai sarana penilaian hasil kemajuan belajar peserta didik,

beberapa di antaranya seperti penilaian yang berupa tes praktik, tes lisan, dan juga penilaian dalam bentuk tes tertulis yang mengharuskan peserta ujian untuk mengisi jawaban dari soal secara tertulis berupa pilihan dan atau uraian (Supriyati dan Dudung, 2018). Penilaian hasil belajar yang berbentuk tes uraian dapat melatih peserta didik dalam menyampaikan pemahaman atas suatu bidang keilmuan secara verbal, yang mana pada tes uraian tidak diberikan pilihan jawaban dan mengharuskan peserta didik untuk menjawab dalam bentuk kalimat sesuai dengan pemahaman yang dimiliki.

Dalam penerapan penilaian secara otomatis, jenis penilaian hasil belajar yang banyak digunakan hanyalah tes tertulis yang berbentuk pilihan ganda, dengan alasan kemudahan dalam proses penilaian tiap butir soalnya, oleh karena itu banyak sekali penelitian yang bertujuan untuk membangun bentuk penilaian yang dapat membantu tenaga pendidikan dalam melakukan penilaian hasil belajar yang berbentuk tes uraian secara otomatis (Hamza dkk., 2013). Salah satu penelitian terkait penilaian otomatis jawaban tes uraian yang paling awal dan cukup berumur panjang karena masih sering dibicarakan hingga saat ini adalah *Project Essay Grade* (PEG) yang dikembangkan oleh Ellis Page pada tahun 1966, setelah itu banyak bermunculan penelitian serupa yang dikembangkan dengan metode lain seperti *Intelligent Essay Assessor* (IEA) yang berbasis *Latent Semantic Analysis* (LSA), kemudian ada *Electronic Essay Rater* (E-Rater) yang menggunakan kombinasi antara teknik statistik dengan *Natural Language Processing* (NLP) untuk mengurai semua kalimat dari jawaban tes uraian yang akan dinilai (Valenti dkk., 2003). Selain dari metode yang disebutkan sebelumnya, terdapat beberapa metode lain yang berbasis pada *string matching* seperti yang dijelaskan oleh Bahri (2014:62) bahwasanya terdapat beberapa algoritma yang dapat digunakan dalam proses *string matching* atau pencocokan *string*, di antaranya seperti algoritma *Knuth Morris Pratt*, *Brute Force*, *Boyer Moore*, dan berikutnya adalah algoritma *Rabin Karp* yang nantinya akan menjadi subjek pembahasan dalam penulisan ini.

Algoritma *Rabin Karp* merupakan algoritma pencarian *string multi pattern* yang menggunakan fungsi *hash* untuk menemukan pola dalam *string* teks, algoritma ini dikembangkan pada tahun 1987 oleh Michael O. Rabin dan Richard M. Karp (Filcha dan Hayaty, 2019). Kemudian dalam penelitian yang dilaksanakan oleh Billhaqqi dkk. (2020) terkait analisis perbandingan antara algoritma *Rabin Karp* dan *Winnowing* dalam penerapannya pada penilaian otomatis jawaban yang berbentuk deskriptif, diperoleh hasil yang menyatakan bahwa algoritma *Rabin Karp* memiliki persentase yang lebih baik dalam melakukan penilaian jawaban otomatis dengan selisih 26% dari penilaian yang dilakukan secara manual, sedangkan

algoritma *Winnowing* mendapatkan selisih yang lebih tinggi yaitu 38% (Billhaqqi dkk., 2020).

Kemudian dalam penelitian terdahulu terkait penilaian otomatis jawaban tes uraian menggunakan algoritma *Rabin Karp* yang dilaksanakan oleh Hamza dkk. (2013) memberikan saran untuk penelitian berikutnya agar menggunakan data uji yang lebih variatif seperti menggabungkan beberapa metode yang sudah ada dengan metode yang lain, juga melakukan penambahan sinonim kata pada sistem sehingga mendapatkan hasil yang lebih baik (Hamza dkk., 2013). Oleh karenanya penulis ingin mencoba untuk menambahkan metode *Synonym Recognition* untuk meningkatkan hasil penilaian yang lebih baik dan lebih akurat. *Synonym Recognition* merupakan proses pencarian sinonim pada setiap kata yang kemudian akan diubah berdasarkan isi kamus sinonim yang tersedia untuk menghindari dua kata berbeda yang memiliki makna serupa, antara jawaban peserta ujian dan kunci jawaban dari tes uraian (Putra dan Sularno, 2019).

Dalam penerapannya nanti, algoritma *Rabin Karp* dengan pendekatan *Synonym Recognition* akan dilakukan setelah melewati tahapan *Text Preprocessing* yang kemudian proses akan dilanjutkan dengan penguraian kalimat atau disebut dengan *Parsing K-Gram*, setelah itu akan dicari nilai *hash* dari setiap kata yang telah diurai dan dilanjutkan dengan menghitung nilai kemiripan antara jawaban peserta ujian dengan kunci jawaban yang tersedia menggunakan metode *Dice Coefficient Similarity*, yang dimana metode tersebut merupakan metode yang paling sering dipakai dalam melakukan perhitungan nilai kemiripan yang berbasis pada pendekatan K-Gram. (Filcha dan Hayaty, 2019; Putra dan Sularno, 2019).

Penelitian ini nantinya akan diterapkan pada mata pelajaran Informatika di Madrasah Aliyah (MA) yang dalam buku Pedoman Implementasi mata pelajaran Informatika K-13 (2019:6) dijelaskan bahwa terdapat beberapa materi pokok atau komponen pengetahuan utama yang akan diberikan kepada peserta didik, yang di antaranya adalah Dampak Sosial Informatika, Jaringan Komputer & Internet, Teknik Komputer, Algoritma & Pemrograman, dan Analisis Data (Balitbang Kemdikbud, 2019:6).

Berdasarkan narasi di atas, maka penulis akan melaksanakan sebuah penelitian yang berjudul “Penerapan Sistem Penilaian Otomatis pada Jawaban Tes Uraian menggunakan Algoritma *Rabin Karp* untuk Mata Pelajaran Informatika di MA Kanjeng Sepuh” yang dalam penelitian ini akan mengembangkan sebuah media penilaian otomatis menggunakan algoritma *Rabin Karp* yang dikombinasikan dengan *Synonym Recognition* untuk mendapatkan hasil penilaian yang lebih akurat.

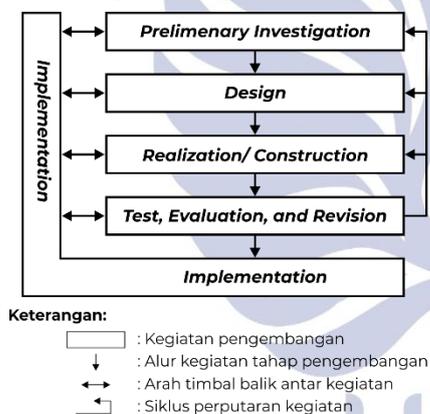
Penelitian ini bertujuan untuk mengetahui cara mengembangkan sistem penilaian otomatis pada jawaban

tes uraian dengan menggunakan algoritma *Rabin Karp*, juga untuk mengetahui hasil uji reliabilitas antara hasil penilaian otomatis yang menggunakan algoritma Rabin Karp dengan hasil penilaian secara manual pada jawaban tes uraian.

## METODE

Dalam penelitian ini digunakan metode *Research and Development* (R&D) dengan menggunakan model penelitian yang dikemukakan oleh Tjeerd Plomp yang selanjutnya dalam penulisan artikel ini akan disebut dengan model Plomp. Metode penelitian model Plomp ini dirancang dengan sesingkat mungkin agar penelitian yang menggunakan model ini dapat berjalan secara sistematis dalam proses pemecahan masalah di dalamnya, khususnya pada penelitian yang didesain untuk bidang pendidikan (Plomp, 1997).

Terdapat beberapa fase dalam penerapan model Plomp, antara lain sebagai berikut: (1) *preliminary investigation* (investigasi); (2) *design* (Desain); (3) *realization/ construction* (realisasi/ konstruksi); (4) *test, evaluation, and revision* (tes, evaluasi, dan revisi); (5) dan *implementation* (implementasi).



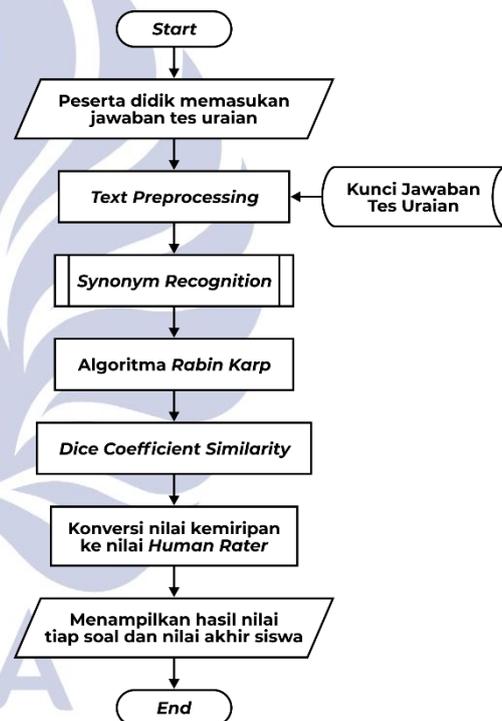
Gambar 1. Desain Metode Penelitian R&D Model Plomp

### Investigasi (*Preliminary Investigation*)

Investigasi awal dilakukan dengan mengkaji penelitian terdahulu terkait penilaian otomatis pada jawaban tes uraian dengan menelaah setiap hasil, simpulan, dan saran pada tiap-tiap penelitian. Kemudian ditemui sebuah penelitian yang dilaksanakan oleh Putra dan Sularno (2019:138) terkait penggunaan algoritma *Rabin Karp* yang dikombinasikan dengan *Synonym Recognition* sebagai pengujian plagiarisme, dan didapati kesimpulan bahwa algoritma *Rabin Karp* yang dikombinasikan dengan *Synonym Recognition* memperoleh hasil yang lebih akurat dibandingkan dengan penerapan algoritma *Rabin Karp* tanpa *Synonym Recognition*. Itulah yang menjadi motivasi penulis dalam mencoba penggunaan algoritma *Rabin Karp* dengan pendekatan *Synonym Recognition* dalam penerapan sistem penilaian otomatis pada jawaban tes uraian.

## Desain (*Design*)

Desain dari rancangan sistem penilaian otomatis jawaban tes uraian menggunakan algoritma *Rabin Karp* dengan pendekatan *Synonym Recognition* akan dimulai dengan proses *Text Preprocessing* pada jawaban peserta ujian beserta kunci jawaban yang telah disediakan oleh guru, kemudian dilanjutkan dengan tahap *Synonym Recognition* yang merupakan proses pencarian sinonim pada setiap kata berdasarkan isi kamus sinonim yang tersedia dalam *database*, selanjutnya proses akan dilanjutkan dengan *Parsing K-Gram* dan *Hasing* yang merupakan tahapan utama dalam algoritma *Rabin Karp*, kemudian nilai kemiripan akan dihitung menggunakan metode *Dice Coefficient Similarity* yang selanjutnya nilai kemiripan akan di konversi ke nilai soal berbentuk *human rates* dengan skala 0-100. Alur proses sistem dapat dilihat pada Gambar 2.



Gambar 2. Flowchart Alur Proses Sistem

Berikut penjelasan terkait tahapan dalam proses sistem penilaian otomatis jawaban tes uraian menggunakan algoritma *Rabin Karp* dengan pendekatan *Synonym Recognition*:

#### a. *Input Data*

Sebelum peserta didik mengerjakan ujian, guru membuat *form* ujian terlebih dahulu dengan memasukkan soal uraian beserta kunci jawabannya, kemudian *form* ujian dibagikan ke peserta ujian untuk dikerjakan, dalam tahapan ini jawaban yang dikirimkan oleh peserta ujian akan disimpan dalam *database*.

b. *Text Preprocessing*

Jawaban tes uraian yang dikirimkan oleh peserta didik akan diproses dalam tahap *Text Preprocessing* yang merupakan sekumpulan prosedur operasi teks yang bertujuan untuk mempermudah proses *text mining* karena mengembalikan tiap-tiap kata dalam sebuah kalimat atau dokumen menjadi bahasa alami atau kata dasar (Katariya dan Chaudhari, 2015). Berikut tahapan umum dalam *Text Preprocessing*:

1) *Case Folding*

Merupakan proses memanipulasi jawaban tes uraian yang dikirimkan oleh peserta didik, yang dimana setiap kata dalam jawaban tes uraian akan diubah menjadi huruf kecil (Filcha dan Hayaty, 2019).

2) *Tokenizing*

Merupakan serangkaian proses pemisahan kata dalam jawaban tes uraian berdasarkan susunan kata, yang kemudian hasil dari proses pemisahan akan disebut dengan *token* (Filcha dan Hayaty, 2019).

3) *Punctuation Removal*

Merupakan serangkaian proses penghapusan karakter-karakter simbol dalam jawaban tes uraian seperti tanda tanya, tanda seru, koma, titik, tanda kutip, dan lain sebagainya (Filcha dan Hayaty, 2019).

4) *Stopword Removal*

Merupakan serangkaian proses menghilangkan atau menghapus kata yang tidak relevan seperti konjungsi dan preposisi dalam jawaban tes uraian (Katariya dan Chaudhari, 2015).

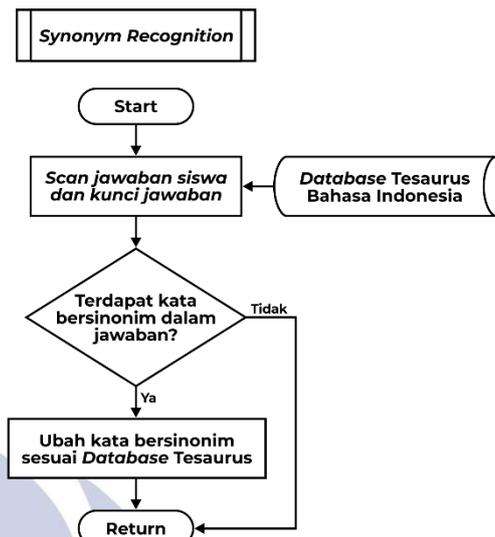
5) *Stemming*

Merupakan proses yang menghasilkan akar kata atau kata dasar dari jamak atau kata kerja seperti 'Bekerja', 'Mengerjakan', dan 'Dikerjakan' nantinya akan direpresentasikan menjadi 'Kerja' (Katariya dan Chaudhari, 2015).

c. *Synonym Recognition*

Jawaban tes uraian yang telah diproses pada tahap *Text Preprocessing* akan dilanjut dengan tahap *Synonym Recognition* yang merupakan proses pencarian sinonim pada setiap kata yang kemudian akan diubah berdasarkan isi kamus sinonim yang tersedia dalam *database* untuk menghindari dua kata berbeda yang memiliki makna serupa, antara jawaban peserta ujian dan kunci jawaban dari soal uraian (Putra dan Sularno, 2019). Alur proses pada

tahap *Synonym Recognition* dapat dilihat pada Gambar 3.



Gambar 3. Flowchart Proses Synonym Recognition

d. *Algoritma Rabin Karp*

Merupakan algoritma pencarian *string multi pattern* yang menggunakan fungsi *hash* untuk menemukan pola dalam *string* teks, algoritma ini dikembangkan pada tahun 1987 oleh Michael O. Rabin dan Richard M. Karp (Filcha dan Hayaty, 2019). Berikut tahapan dalam algoritma *Rabin Karp*:

1) *Parsing K-Gram*

Merupakan serangkaian *terms* dengan panjang K, dan *terms* yang digunakan biasanya berbentuk teks, *Parsing K-Gram* diterapkan untuk memotong karakter huruf sejumlah K dari sebuah kalimat atau kata secara terus menerus hingga keseluruhan kalimat telah terurai (Putra dan Sularno, 2019).

2) *Hashing*

Setelah jawaban tes uraian peserta didik diurai melalui proses *Parsing K-Gram*, selanjutnya akan masuk dalam tahap *Hashing* yang merupakan proses konversi karakter *string* menjadi integer yang nantinya *output* dari proses ini adalah nilai *hash* (Filcha dan Hayaty, 2019).

e. *Dice Coefficient Similarity*

Merupakan metode yang paling sering dipakai dalam melakukan perhitungan nilai kemiripan yang berbasis pada pendekatan K-Gram. (Filcha dan Hayaty, 2019). Persamaan *Dice Coefficient Similarity* dapat dilihat pada Persamaan (1).

$$S = \frac{2 * C}{A + B} \quad (1)$$

Pada Persamaan (1),  $S$  merupakan nilai kemiripan atau *similarity*,  $A$  merupakan nilai Hash dari jawaban tes uraian dan  $B$  merupakan nilai Hash dari kunci jawaban tes uraian, kemudian  $C$  merupakan jumlah nilai Hash yang sama antara jawaban tes uraian dengan kunci jawaban.

f. Konversi nilai kemiripan

Nilai kemiripan yang telah dihitung sebelumnya, akan dikonversi ke nilai soal berbentuk *human rates* dengan skala 0-100, dengan rentang nilai yang digunakan mengacu pada penelitian yang dilaksanakan oleh Fitri dan Asyikin (2015:91). Pada Tabel 1. di bawah ini akan ditampilkan rentang nilai antaran nilai kemiripan dan nilai *human rates*.

Tabel 1. Perbandingan Rentang Nilai

Nilai Kemiripan	Nilai Human Rates
0.01 – 0.10	10 (Sepuluh)
0.11 – 0.20	20 (Dua Puluh)
0.21 – 0.30	30 (Tiga Puluh)
0.31 – 0.40	40 (Empat Puluh)
0.41 – 0.50	50 (Lima Puluh)
0.51 – 0.60	60 (Enam Puluh)
0.61 – 0.70	70 (Tujuh Puluh)
0.71 – 0.80	80 (Delapan Puluh)
0.81 – 0.90	90 (Sembilan Puluh)
0.91 – 1.00	100 (Seratus)

(Fitri dan Asyikin, 2015:91)

**Realisasi/ Konstruksi (Realization/ Construction)**

Sistem yang akan dikembangkan berbasis web dengan menggunakan beberapa teknologi di dalamnya seperti Laravel 8 yang merupakan *framework* PHP, Bootstrap 4 sebagai *framework* CSS, dan jQuery yang merupakan *Library* JavaScript dengan keistimewaan dan kesederhanaan sintak yang tidak terlalu rumit dalam penulisan kodenya.

Algoritma *Rabin Karp* dalam sistem penilaian otomatis dikembangkan dengan menggunakan bahasa Python dan memakai *Library* NLTK (*Natural Language Toolkit*) untuk melakukan *Text Preprocessing*, juga menggunakan *Library* Sastrawi untuk mengembalikan tiap kata dalam kalimat menjadi kata dasar dengan cara membuang imbuhan seperti awalan, akhiran dan atau sisipan, kemudian ditambahkan juga Tesaurus Bahasa Indonesia ke *database* sebagai penunjang proses *Synonym Recognition*.

**Tes, Evaluasi dan Revisi**

Sistem yang dikembangkan akan dilakukan uji validasi oleh beberapa dosen dan tenaga pendidik yang ahli di bidang teknologi dalam dunia pendidikan, dan akan

dilakukan evaluasi serta revisi hingga ditemui titik kelayakan dari sistem yang dikembangkan. Seperti yang dijelaskan oleh Plomp (1992) bahwa penelitian yang dilakukan tanpa adanya evaluasi akan sulit untuk dinyatakan apakah suatu permasalahan dalam rumusan masalah telah diselesaikan sesuai dengan yang diharapkan atau tidak.

Dilakukan juga uji reliabilitas dan uji kesepakatan antara kedua hasil nilai menggunakan koefisien *Cohen’s Kappa* dan koefisien *intra-class correlation* (ICC) dengan bantuan aplikasi IBM SPSS Statistics 25. Koefisien *Kappa* merupakan indeks *Inter-Rater Agreement* yang paling umum digunakan dalam pengukuran kesepakatan terhadap penilaian variabel kategoris yang dinilai oleh dua penilai atau *Rater*, sedangkan *intra-class correlation* (ICC) secara umum digunakan untuk mengukur kesepakatan dan reliabilitas ketika terdapat banyak kategori penilaian atau jika penilaian menggunakan skala berkelanjutan (Shweta dkk., 2015).

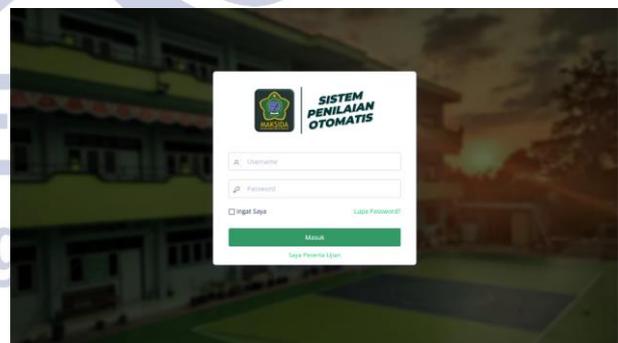
**Implementasi (Implementation)**

Setelah melalui proses validasi dan kemudian sistem penilaian otomatis yang dikembangkan telah dinyatakan valid, maka selanjutnya sistem penilaian otomatis jawaban tes uraian dapat diimplementasikan dalam proses penilaian hasil belajar peserta didik.

**HASIL DAN PEMBAHASAN**

**Implementasi Media Penilaian Otomatis**

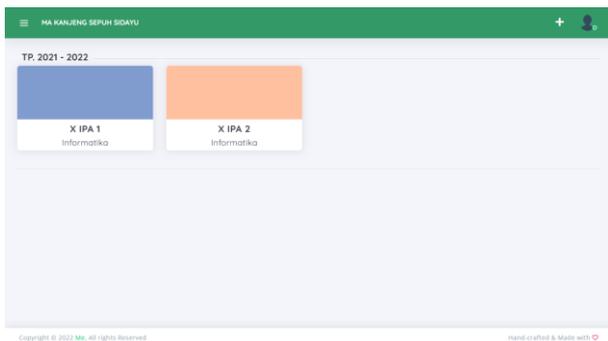
Media penilaian otomatis yang dikembangkan berbasis web dan dapat diakses secara *online* melalui *link* yang tersedia.



Gambar 4. Halaman Login

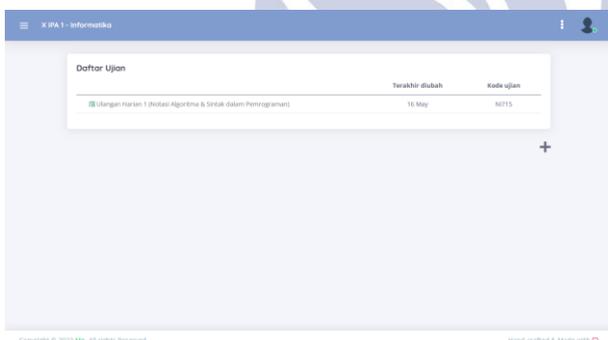
Gambar 4. merupakan halaman yang pertama kali muncul ketika media penilaian otomatis diakses melalui *link* yang tersedia, pada halaman *login* ini *User* diminta untuk memasukkan *Password* dan *Username*, sebelum masuk ke dalam sistem, yang kemudian akan diarahkan ke *dashboard* awal atau halaman utama seperti yang terlihat pada Gambar 5. dan akan menampilkan kelas-kelas yang telah dibuat oleh guru, lengkap dengan tahun pelajaran, nama kelas, serta mata pelajaran yang diampu, setiap kelas

memiliki warna yang dapat diatur sesuai dengan kebutuhan guru agar dapat dengan mudah membedakan satu kelas dengan kelas yang lainnya secara visual. Juga terdapat tombol '+' untuk menambah kelas berikutnya.



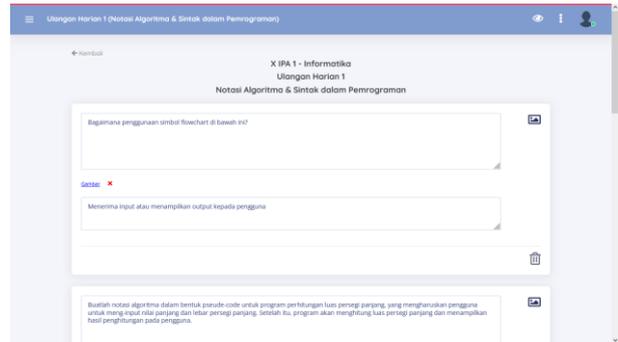
Gambar 5. Halaman Utama

Setelah memilih dari salah satu kelas yang tersedia, guru akan diarahkan ke halaman kelas, seperti yang terlihat pada Gambar 6. dan akan menampilkan form ujian yang telah dibuat, serta dapat menambahkan ujian berikutnya dengan menekan tombol '+' di bawah tabel daftar ujian. Selain itu juga terdapat opsi pengaturan kelas seperti edit kelas dan hapus kelas pada menu ⋮ di pojok kanan atas



Gambar 6. Halaman Kelas

Setelah membuat dan memilih form ujian pada halaman kelas, maka guru akan masuk ke halaman form ujian seperti yang terlihat pada Gambar 7. untuk melakukan penataan pada form ujian, mulai dari menuliskan soal uraian hingga mengisi kunci jawaban. Kemudian pada pojok kanan atas terdapat opsi view yang disimbolkan dengan mata untuk melihat tampilan form ujian yang akan dikerjakan oleh peserta didik nantinya, juga terdapat opsi pengaturan form pada menu ⋮ seperti edit ujian, hapus ujian, aktifkan dan non-aktifkan ujian, pengacakan urutan soal, dan juga tombol untuk melihat hasil jawaban ujian yang telah dikerjakan oleh peserta didik.



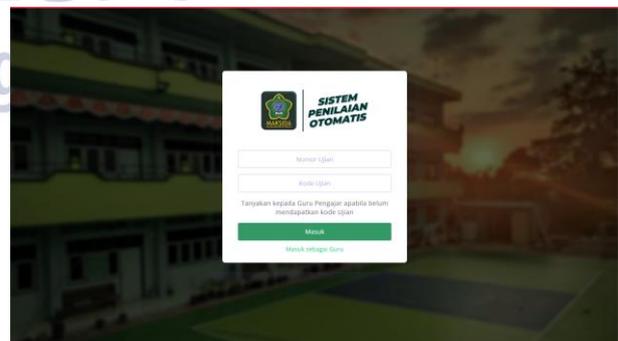
Gambar 7. Halaman Form Ujian

Jawaban peserta ujian yang telah dikirimkan dapat dilihat pada halaman hasil jawaban, dengan memilih opsi "Hasil Jawaban" pada halaman form ujian. Untuk siswa yang sudah mengerjakan dan mengirimkan jawaban maka nilai akan secara otomatis terisi di dalam tabel, dan guru dapat melihat jawaban yang dikirimkan dengan menekan nama peserta didik, guru juga bisa menghapus jawaban peserta didik jika diperlukan. Tampilan halaman hasil jawaban dapat dilihat pada Gambar 8.



Gambar 8. Halaman Hasil Jawaban

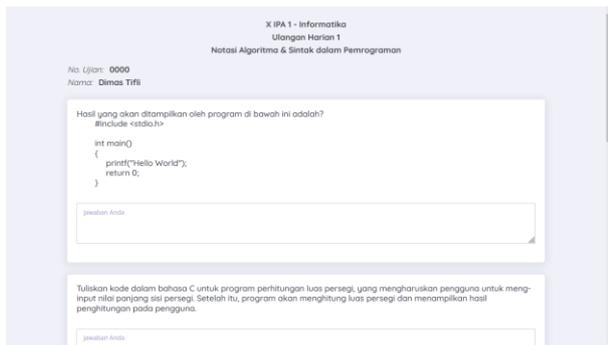
Selanjutnya untuk peserta didik yang ingin mengerjakan ujian diminta memasukkan Nomor Ujian dan Kode Ujian yang diberikan oleh guru terlebih dahulu. Tampilan halaman masuk peserta ujian dapat dilihat pada Gambar 9.



Gambar 9. Halaman Masuk Peserta Ujian

Setelah memasukkan Nomor Ujian dan Kode Ujian yang diberikan oleh guru, peserta didik akan masuk ke dalam form ujian sesuai dengan kode ujian yang diberikan

oleh guru sebelumnya. Tampilan *form* ujian yang akan dikerjakan oleh peserta ujian dapat dilihat pada Gambar 10.



Gambar 10. Form Ujian pada Tampilan Peserta Didik

### Hasil Validasi Instrumen

Perangkat instrumen dalam penelitian ini telah divalidasi oleh validator sejumlah tiga orang, yakni dua Dosen Jurusan Teknik Informatika Universitas Negeri Surabaya dan salah seorang Guru Madrasah Aliyah Kanjeng Sepuh Sidayu. Dalam proses validasi menggunakan skala *likert* dengan menggunakan kriteria skor persentase nilai validasi yang tertera pada Tabel 2.

Tabel 2. Persentase Nilai Validasi

Kriteria	Persentase (%)
Tidak Valid	0 - 20
Kurang Valid	21 - 40
Cukup Valid	41 - 60
Valid	61 - 80
Sangat Valid	81 - 100

(Riduwan, 2015)

Rencana Pelaksanaan Pembelajaran (RPP) merupakan instrumen penelitian pertama yang divalidasi, dalam lembar validasi RPP berisikan terkait tujuan pembelajaran, langkah-langkah pembelajaran, dan teknik penilaian dalam pembelajaran. Lembar validasi soal uraian berisikan terkait validasi isi soal uraian, serta bahasa dan penulisan soal uraian. Lembar validasi media penilaian otomatis berisikan terkait kualitas kegunaan, kualitas informasi, kualitas interaksi pelayanan, dan kepuasan pengguna. Kemudian hasil validasi telah terpapar pada Tabel 3.

Tabel 3. Hasil Validasi Instrumen

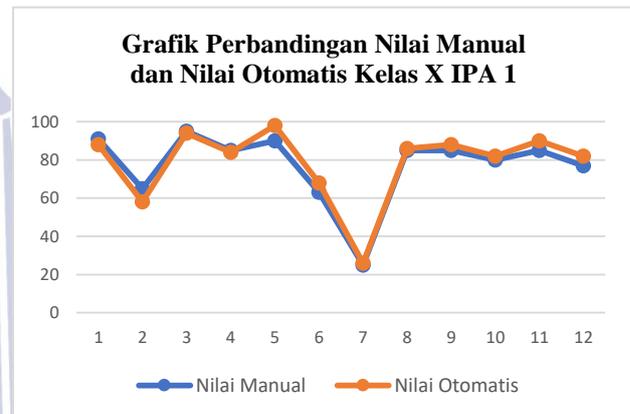
No.	Validasi	Rata-rata Persentase (%)	Kategori
1	RPP	91,50	Sangat Valid
2	Soal	90,56	Sangat Valid
3	Media	90,51	Sangat Valid

### Hasil Pengujian Media Penilaian Otomatis

Penelitian dilaksanakan di Madrasah Aliyah Kanjeng Sepuh Sidayu, yang berlokasi di Jl. Pemuda, Bunderan, Kec. Sidayu, Kabupaten Gresik, Jawa Timur 61153. Pengujian diterapkan pada mata pelajaran Informatika

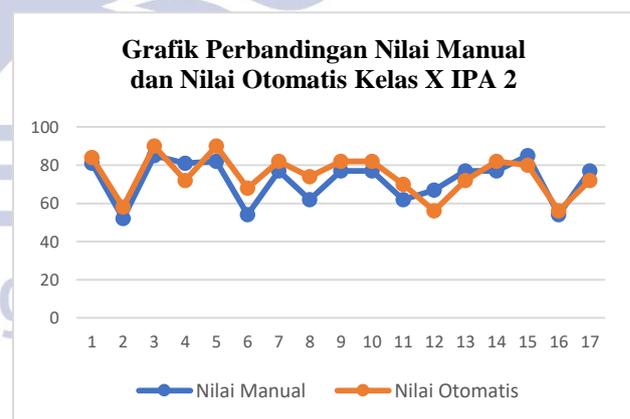
kelas X dengan jumlah soal yang diujikan sebanyak 5 butir soal uraian pada materi Algoritma & Pemrograman, dengan jumlah peserta ujian yang dibagi menjadi tiga kelas, yaitu kelas X IPA 1, X IPA 2, dan X IPS 1, dengan total peserta ujian sebanyak 55 siswa.

Pada kelas X IPA 1 dengan jumlah peserta ujian sebanyak 12 siswa, didapati hasil rata-rata nilai manual sebesar 77,17 dan rata-rata nilai secara otomatis sebesar 78,67. Perbandingan hasil perhitungan nilai secara manual dengan nilai secara otomatis pada kelas X IPA 1 dapat dilihat dalam Gambar 11.



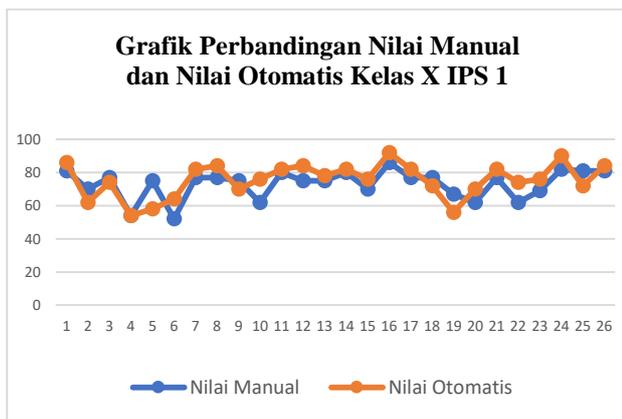
Gambar 11. Perbandingan Nilai Kelas X IPA 1

Pada kelas X IPA 2 dengan jumlah peserta ujian sebanyak 17 siswa, didapati hasil rata-rata nilai manual sebesar 72,18 dan rata-rata nilai secara otomatis sebesar 74,71. Perbandingan hasil perhitungan nilai secara manual dengan nilai secara otomatis pada kelas X IPA 2 dapat dilihat dalam Gambar 12.



Gambar 12. Perbandingan Nilai Kelas X IPA 2

Pada kelas X IPS 1 dengan jumlah peserta ujian sebanyak 26 siswa, didapati hasil rata-rata nilai manual sebesar 73,12 dan rata-rata nilai secara otomatis sebesar 75,46. Perbandingan hasil perhitungan nilai secara manual dengan nilai secara otomatis pada kelas X IPS 1 dapat dilihat dalam Gambar 13.



Gambar 13. Perbandingan Nilai Kelas X IPS 1

Dapat dilihat dari masing-masing kelas terdapat selisih rata-rata yang kurang dari 2,6 dan secara keseluruhan dengan jumlah peserta ujian sebanyak 55 siswa didapati hasil rata-rata nilai manual sebesar 74,15 dan rata-rata nilai secara otomatis sebesar 76,28.

Perhitungan secara statistik untuk uji reliabilitas dan uji kesepakatan antara kedua hasil nilai akan dihitung menggunakan koefisien *Cohen's Kappa* dan koefisien *intra-class correlation* (ICC) pada masing-masing butir soal. Hasil perhitungan uji reliabilitas dan uji kesepakatan dapat dilihat pada Tabel 4.

Tabel 4. Hasil Uji Reliabilitas & Uji Kesepakatan

No. Soal	Kappa	ICC
1	0.179	0.910
2	0.141	0.758
3	0.128	0.911
4	0.483	0.792
5	-0.001	0.384

Bentuk soal dalam penelitian ini terdiri dari 4 soal uraian singkat dan 1 soal yang mengharuskan peserta ujian untuk menuliskan sebuah fungsi kode pemrograman yaitu pada soal nomor 5. Kemudian didapati hasil perhitungan menggunakan koefisien *Kappa* pada Tabel 4 menunjukkan bahwa soal nomor 5 memperoleh hasil kategori *Poor* karena didapati nilai di bawah nol ( $<0.00$ ), kemudian didapati pula hasil kategori *Moderate* ( $>0.40$ ) pada nomor 4, dan untuk ketiga butir soal lainnya memperoleh kategori *Slight* ( $>0.10$ ). Secara keseluruhan hasil uji kesepakatan antar kedua *Rater* menggunakan koefisien *Kappa* terhitung relatif rendah. Rendahnya hasil uji tersebut diakibatkan dari hasil nilai yang tidak sama persis antara penilaian otomatis dengan penilaian secara manual.

Hasil perhitungan menggunakan ICC pada Tabel 4 menunjukkan bahwa 4 dari 5 butir soal memperoleh hasil kategori *Good* ( $>0.70$ ) dan salah satu butir soal yang lain memperoleh hasil kategori *Bad* ( $<0.40$ ) yaitu pada soal nomor 5. Secara keseluruhan hasil uji reliabilitas dan uji kesepakatan antar kedua *Rater* menggunakan ICC

terhitung relatif tinggi. Hasil uji yang relatif tinggi menggunakan ICC ini menunjukkan bahwa kedua *Rater* yang diujikan memiliki tingkat reliabilitas penilaian yang relatif sama antar keduanya.

Hasil pengujian kesepakatan dan reliabilitas yang relatif rendah saat menggunakan koefisien *Kappa* dan hasil yang relatif tinggi saat menggunakan ICC, maka dapat diindikasikan bahwa hasil penilaian secara manual dan penilaian secara otomatis menggunakan algoritma *Rabin Karp* dengan pendekatan *Synonym Recognition* ini masuk ke dalam kategori *Low Agreement with High Reliability* atau tingkat kesepakatan yang rendah namun dengan reliabilitas yang tinggi, seperti yang dijelaskan oleh Shweta dkk. (2015) bahwa pengukuran kesepakatan itu digunakan untuk menunjukkan seberapa sering dua *Rater* menetapkan peringkat yang sama persis, sedangkan reliabilitas digunakan untuk mengukur kesamaan yang relatif antara kedua *Rater*, sehingga dua *Rater* yang memiliki kesepakatan rendah masih dapat memiliki reliabilitas yang tinggi.

Permana dkk. (2021) juga menjelaskan bahwa reliabilitas dan kesepakatan yang tinggi antar *Rater* tentunya sangat diharapkan dalam setiap penelitian terkait penilaian soal uraian secara otomatis, namun reliabilitas penilaian antara manusia dan mesin tidak akan pernah sekuat reliabilitas antara penilai manusia dan manusia, oleh karenanya untuk mencapai nilai reliabilitas yang sangat tinggi dalam penilaian soal uraian otomatis merupakan suatu permasalahan yang cukup sulit. Oleh karena itu dibutuhkan pemilihan bentuk soal yang tepat serta penentuan rubrik penilaian dan konversi nilai otomatis yang cocok antar keduanya.

## PENUTUP

### Simpulan

Berikut simpulan yang dapat diambil dalam penelitian ini, antara lain:

1. Penerapan Sistem Penilaian Otomatis pada Jawaban Tes Uraian menggunakan Algoritma Rabin Karp untuk Mata Pelajaran Informatika di MA Kanjeng Sepuh telah dilaksanakan sesuai prosedur pengembangan menggunakan metode R&D model Plomp
2. Seluruh perangkat dalam penelitian telah divalidasi oleh 3 validator ahli dan memperoleh hasil persentase 90,51% untuk media penilaian otomatis jawaban tes uraian, kemudian persentase 91,50% untuk Rencana Pelaksanaan Pembelajaran (RPP), dan terakhir untuk soal uraian diperoleh hasil persentase sebesar 90,56% yang dimana ketiga instrumen penelitian tersebut mendapat kategori sangat valid.

3. Sistem penilaian otomatis pada jawaban tes uraian menggunakan algoritma Rabin Karp telah diujikan kepada 55 siswa yang terbagi menjadi 3 kelas, dan didapati hasil rata-rata nilai manual sebesar 74,15 dan rata-rata nilai secara otomatis sebesar 76,28.
4. Hasil uji reliabilitas dan uji kesepakatan menggunakan koefisien *Kappa* pada tiap butir soal didapati hasil perhitungan yang relatif rendah, namun untuk perhitungan uji reliabilitas dan uji kesepakatan menggunakan ICC didapati hasil yang relatif tinggi, hal tersebut dapat diindikasikan bahwa hasil penilaian secara manual dan penilaian secara otomatis menggunakan algoritma *Rabin Karp* dengan pendekatan *Synonym Recognition* ini masuk ke dalam kategori *Low Agreement with High Reliability* atau tingkat kesepakatan yang rendah namun dengan reliabilitas yang tinggi.

#### Saran

Berikut beberapa poin yang dapat penulis sarankan untuk penelitian selanjutnya:

1. Pemilihan bentuk soal yang tepat serta penentuan rubrik penilaian dan konversi nilai otomatis yang cocok antar keduanya.
2. Penelitian yang telah dilaksanakan ini hanya berfokus pada mata pelajaran Informatika, untuk itu diharapkan penelitian selanjutnya dapat menerapkan sistem penilaian otomatis jawaban tes uraian menggunakan algoritma *Rabin Karp* dengan pendekatan *Synonym Recognition* ini pada mata pelajaran yang lain.

#### DAFTAR PUSTAKA

- Badan Penelitian dan Pengembangan. (2019). *Pedoman Implementasi Muatan/Mata Pelajaran Informatika Kurikulum 2013*. Jakarta: Balitbang Kemdikbud
- Bahri, S. (2014). Penilaian Otomatis Ujian Essay Online berbasis Algoritma Rabin Karp. *SWABUMI, Vol. 1*.
- Billhaqqi, T. T. I., Wicaksono, G. W., Aditya, C. S. K. (2020). Analisis Perbandingan Algoritma Rabin Karp dan Winnowing dalam Penelitian Jawaban Otomatis. *Prosiding SENTRA (Seminar Teknologi dan Rekayasa)*, No. 6, 269-276.
- Filcha, A., & Hayaty, M. (2019). Implementasi Algoritma Rabin-Karp untuk Pendeteksi Plagiarisme pada Dokumen Tugas Mahasiswa. *JUITA: Jurnal Informatika*, 7(1), 25-32.
- Fitri, Rahimi, & Asyikin, Arifin Noor. (2015). Aplikasi penilaian ujian essay otomatis menggunakan metode cosine similarity. *Poros Teknik*, 7(2), 88-94.
- Hamza, S., Sarosa M., & Santoso, P. B. (2013). Sistem Koreksi Soal Essay Otomatis dengan Menggunakan Metode Rabin Karp. *Jurnal EECCIS Vol. 7*, No. 2, 153-158.
- Katariya, N. P., & Chaudhari, M. S. (2015). Text preprocessing for text mining using side information. *Int. J. Comput. Sci. Mob. Appl*, Vol. 3, 3-7.
- Permana, R. A. H. A., Widodo, A., Setiawan, W., & Sriyati, S. (2021). Penggunaan Penskor Jawaban Esai Otomatis dalam Pengukuran Pengetahuan Guru. *Jurnal IPA & Pembelajaran IPA*, 5(4), 279-292.
- Plomp, T. (1997). Educational Design: Introduction. From Tjeerd Plomp (Eds.). *Educational & Training System Design: Introduction. Design of Education and Training (in Dutch)*. Utrecht: Lemma. Netherland. Faculty of Educational Science and Technology, University of Twente.
- Plomp, T., & van den Wolde, J. (1992). The General Model for Systematical Problem Solving. From Tjeerd Plomp (Eds.). *Design of educational and training (in Dutch)*. Utrecht: Lemma. Netherland. Faculty of Educational Science and Technology, University of Twente.
- Putra, N. P., & Sularno, S. (2019). Penerapan Algoritma Rabin-karp dengan Pendekatan Synonym Recognition sebagai Antisipasi Plagiarisme pada Penulisan Skripsi. *Jurnal Teknologi Dan Sistem Informasi Bisnis-JTEKSIS*, 1(2), 130-140.
- Riduwan, M. B. A. (2015). *Dasar-dasar statistika*. Bandung: Alfabeta.
- Shweta, Bajpai, R. C., & Chaturvedi, H. K. (2015). Evaluation of inter-rater agreement and inter-rater reliability for observational data: an overview of concepts and methods. *Journal of the Indian Academy of Applied Psychology*, 41(3), 20-27.
- Supriyati, Y. & Dudung A. (2018). *Penilaian Kelas*. Depok: Karya Ilmu Media Aulia.
- Valenti, S., Neri, F., & Cucchiarelli, A. (2003). An overview of current research on automated essay grading. *Journal of Information Technology Education: Research*, 2(1), 319-330.