



# Sistem Pengukuran Tingkat Kemiripan Judul dan Abstrak Skripsi Menggunakan Algoritma WInnowing dan Dice Similarity

Hadi Sucipto<sup>1\*</sup>, Aries Dwi Indriyanti<sup>2</sup>, Kartika Mauludi<sup>1</sup>, Ahmad Heru Mujianto<sup>2</sup>, Muhammad Fatkhur Rizal<sup>1</sup>, Chamdan Mashuri<sup>2</sup>

<sup>1</sup>Fakultas Teknologi Informasi, Teknik Informatika, Universitas Hasyim Asy'ari, Jombang Tebuireng, Jl. Irian Jaya No.55, Cukir, Kec. Diwek, Kabupaten Jombang, Jawa Timur, Indonesia

<sup>2</sup>Fakultas Teknologi Informasi, Sistem Informasi, Universitas Hasyim Asy'ari, Jombang Tebuireng, Jl. Irian Jaya No.55, Cukir, Kec. Diwek, Kabupaten Jombang, Jawa Timur, Indonesia

Email: <sup>1,\*</sup>hadisucipto@unhasy.ac.id, <sup>2</sup>ariesdwiindriyanti@unhasy.ac.id, <sup>3</sup>kartikamauludi@mhs.unhasy.ac.id,

<sup>4</sup>ahmadmujianto@unhasy.ac.id, <sup>5</sup>fatkhurrizal@unhasy.ac.id, <sup>6</sup>chamdanmashuri@unhasy.ac.id

Email Penulis Korespondensi: hadisucipto@unhasy.ac.id

Submitted: 22/09/2024; Accepted: 10/10/2024; Published: 19/10/2024

**Abstrak**—Skripsi merupakan karya tulis ilmiah yang wajib dibuat oleh akademisi tingkat strata 1 di fakultas teknologi informasi UNHASY. Salah satu masalah dari pembuatan skripsi di fakultas teknologi informasi UNHASY yaitu masih manualnya proses cek kemiripan judul dan abstrak sehingga membutuhkan waktu yang lama dalam proses pengecekannya. Dari permasalahan yang ada penelitian dilakukan dengan bertujuan untuk menyediakan sebuah sistem Pengukuran Tingkat Kemiripan Judul Dan Abstrak Skripsi Berbasis Web Menggunakan Algoritma WInnowing Dan Dice Similarity. Algoritma yang digunakan pada penelitian ini adalah algoritma winnowing yang digunakan sebagai document fingerprinting. Judul dan abstrak skripsi yang di masukan akan diolah melalui beberapa tahapan yaitu Whitespace insensitive, membentuk rangkaian n-gram, rolling hash, pemecahan nilai hash menjadi beberapa window kemudian pengambilan fingerprint terkecil dari setiap window. Pada tahap akhir dicari nilai similarity dengan menggunakan dice coefficient. Hasil dari penelitian ini merupakan sebuah sistem yang dapat mengecek kemiripan judul dan abstrak dengan menggunakan algoritma winnowing. Algoritma ini tidak memiliki ketentuan khusus mengenai berapa nilai gram, window, dan prima yang harus diterapkan. Sehingga penulis melakukan pengujian dengan menggunakan 10 data uji untuk mengetahui pengaruh nilai tersebut. Sehingga diperoleh nilai terkecil yang diperoleh dari nilai masukan gram = 5, window = 4 dan prima = 3 yaitu 10.82%, sedangkan hasil terbesar yaitu 38.16% diperoleh dari nilai masukan gram = 4, window = 3 dan prima = 3.

**Kata Kunci:** WInnowing; Dice Coefficient; Plagiarisme; Algoritma WInnowing; Sistem Pengukuran Kemiripan

**Abstract**—Thesis is a scientific paper that must be made by undergraduate stratum 1 academics at the UNHASY information technology faculty. One of the problems with writing a thesis at the Faculty of Information Technology, UNHASY, is that the process of checking the similarity of titles and abstracts is still manual so it takes a long time to check. From the existing problems, the research was carried out with the aim of providing a system for measuring the level of similarity of titles and abstracts of web-based thesis using the WInnowing and Dice Similarity Algorithm. The algorithm used in this research is the winnowing algorithm which is used as document fingerprinting. The title and abstract of the submitted thesis will be processed through several stages, namely Whitespace insensitive, forming an n-gram sequence, rolling hash, splitting the hash value into several windows and then taking the smallest fingerprint from each window. In the final stage, the similarity value is sought by using the dice coefficient. The results of this study are a sub-system that can check the similarity of titles and abstracts using the winnowing algorithm. This algorithm does not have specific provisions regarding how many gram, window, and prime values that must be applied so that the author tests using 10 test data to determine the effect of these values so that the smallest value is obtained from the input value gram = 5, window = 4 and prime = 3, which is 10.82%, while the largest result, which is 38.16%, is obtained from the input value of gram = 4, window = 3 and prime = 3.

**Keywords:** WInnowing; Dice Coefficient; Plagiarism; WInnowing Algorithm; Similarity Measurement System

## 1. PENDAHULUAN

Dalam konteks akademik, penyusunan skripsi merupakan tahap penting bagi mahasiswa Strata Satu (S1) sebagai bagian dari persyaratan untuk memperoleh gelar sarjana. Skripsi tidak hanya berfungsi sebagai tugas akhir yang menunjukkan kemampuan mahasiswa dalam melakukan penelitian, tetapi juga sebagai kontribusi terhadap solusi permasalahan nyata yang dihadapi masyarakat. Oleh karena itu, relevansi topik penelitian yang diangkat dalam skripsi sangat penting untuk diperhatikan agar memiliki dampak yang lebih besar, baik secara akademik maupun praktis [1] [2]. Namun, seiring dengan meningkatnya jumlah mahasiswa yang menyelesaikan skripsi setiap tahunnya, muncul permasalahan baru yaitu plagiarisme.

Dalam banyak kasus, risiko plagiarisme sering kali disebabkan oleh kesulitan dalam memeriksa keunikan judul skripsi yang diajukan. Di banyak institusi pendidikan, pengajuan judul skripsi masih dilakukan secara manual, tanpa adanya sistem otomatis yang mampu mengecek kesamaan judul dengan judul skripsi yang sudah ada sebelumnya [3] [4] [5]. Hal ini menyebabkan banyak mahasiswa yang secara tidak sadar mengajukan judul skripsi yang sudah digunakan oleh mahasiswa lain, meningkatkan potensi plagiarisme dan menurunkan nilai orisinalitas dari karya ilmiah yang dihasilkan.

Penelitian menunjukkan bahwa kurangnya kesadaran akan pentingnya orisinalitas dalam karya akademik dapat memperparah perilaku plagiarisme di kalangan mahasiswa [6] Mahasiswa sering kali tidak memahami betul



bahwa plagiarisme tidak hanya merugikan mereka secara akademik tetapi juga menghambat proses penyebaran ilmu pengetahuan yang kredibel dan valid[7].

Selain itu, dosen pembimbing, meskipun memiliki peran penting dalam proses bimbingan skripsi, sering kali kesulitan untuk memeriksa secara manual keunikan dari setiap judul skripsi yang diajukan oleh mahasiswa yang mereka bimbing[8] [9]. Untuk mengatasi masalah plagiarisme ini, diperlukan sebuah sistem yang dapat memeriksa kemiripan judul dan abstrak skripsi secara otomatis. Sistem ini harus mampu mengidentifikasi kesamaan antara judul atau abstrak yang diajukan dengan skripsi yang sudah ada di database kampus. Dengan demikian, mahasiswa dapat segera mengetahui jika judul yang mereka ajukan memiliki tingkat kesamaan yang tinggi dengan skripsi sebelumnya, dan mereka dapat melakukan penyesuaian sebelum melanjutkan ke tahap penelitian lebih lanjut.

Pada penelitian ini, algoritma WInnowing dan metode Dice Similarity digunakan sebagai metode untuk mengukur tingkat kemiripan antara judul dan abstrak skripsi. Algoritma WInnowing adalah salah satu algoritma yang digunakan dalam pengecekan kemiripan teks dan memiliki keunggulan dalam deteksi kemiripan yang lebih halus dengan cara melakukan "document fingerprinting" [10]. Sedangkan Dice Similarity adalah metode yang umum digunakan untuk membandingkan dua set data teks dan menghasilkan skor kemiripan berdasarkan rasio overlap antar n-gram dalam teks yang dibandingkan [11][12]. Metode WInnowing dan Dice Similarity dipilih karena mereka terbukti efektif dalam mendeteksi kesamaan teks di berbagai aplikasi sebelumnya, dan penelitian ini mencoba menerapkan metode tersebut dalam konteks pengecekan judul dan abstrak skripsi.

Sistem semacam ini telah banyak diteliti dalam beberapa tahun terakhir. Salah satu penelitian yang relevan dilakukan oleh Cahyapratama [11] dalam karya mereka yang berjudul "Gap analysis business process model by using structural similarity". Penelitian tersebut menggunakan metode Dice Coefficient untuk membandingkan similaritas dua teks dan menunjukkan bahwa metode ini efektif dalam deteksi kemiripan teks dengan akurasi sebesar 36,25% menggunakan K-Fold Cross Validation pada 1000 data Tugas Akhir [11]. Selain itu, penggunaan algoritma WInnowing dan persamaan Jaccard similarity juga telah diterapkan dalam berbagai penelitian untuk mendeteksi kemiripan teks. Jaccard similarity adalah metode yang mengukur kemiripan antara dua set data dengan membandingkan jumlah data yang sama di kedua set tersebut. Penelitian Kamaruddin juga menunjukkan bahwa Jaccard similarity mampu memberikan hasil yang baik dalam konteks pengolahan bahasa alami dan analisis teks [13] [14].

Penelitian mengenai sistem pengukuran tingkat kemiripan judul dan abstrak skripsi menggunakan algoritma WInnowing dan Dice Similarity telah berkembang secara signifikan dalam beberapa tahun terakhir, terutama terkait dengan deteksi plagiarisme dan analisis kesamaan teks. Penelitian yang dilakukan oleh Purwaningrum [15] menyoroti penggunaan algoritma WInnowing yang digabungkan dengan Dice Similarity dan Jaccard Coefficient untuk mendeteksi kemiripan teks dalam dokumen berbahasa Indonesia. Hasil penelitian menunjukkan bahwa kombinasi algoritma WInnowing dengan Dice Similarity memberikan akurasi terbaik dengan nilai kesamaan rata-rata 71,18%, dibandingkan dengan Jaccard yang hanya mencapai 35,59% .

Lebih lanjut, penelitian oleh Sunardi menambahkan elemen pengenalan sinonim ke dalam algoritma WInnowing, yang secara signifikan meningkatkan akurasi deteksi kesamaan teks. Penelitian ini menunjukkan bahwa pengenalan sinonim dalam proses deteksi kesamaan teks meningkatkan akurasi hingga 3,11%, sehingga memperkuat efektivitas algoritma WInnowing dalam menangani variasi bahasa dalam dokumen akademik [16]. Penelitian Soleh & Purwarianti lebih fokus pada pengembangan sistem deteksi kesamaan teks menggunakan kombinasi algoritma WInnowing dan metode N-Gram. Penelitian ini menemukan bahwa kombinasi tersebut mampu memberikan hasil deteksi yang lebih akurat dalam identifikasi plagiarisme teks dibandingkan dengan metode tradisional lainnya, terutama dalam mendeteksi frasa-frasa mirip di dalam dokumen [17].

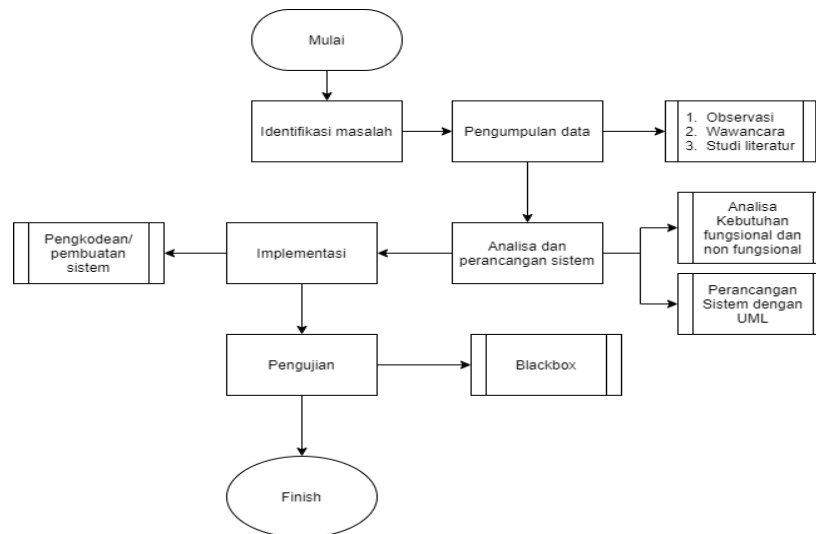
Terakhir, penelitian oleh Alamsyah mengaplikasikan algoritma WInnowing untuk mendeteksi kemiripan judul penelitian guna menghindari duplikasi dalam pengajuan proposal penelitian. Hasil penelitian menunjukkan bahwa algoritma ini sangat efektif dalam mengidentifikasi judul-judul yang mirip, sehingga dapat mencegah plagiarisme dalam proses penulisan akademik [18].

Kombinasi dari algoritma WInnowing dan Dice Similarity dalam berbagai penelitian ini secara konsisten menunjukkan hasil yang signifikan dalam meningkatkan akurasi deteksi kesamaan teks, baik untuk keperluan plagiarisme maupun dalam pengukuran kesamaan judul dan abstrak dalam konteks akademik. Perbedaan signifikan dari penelitian ini adalah bahwa metode yang dikembangkan lebih fokus pada sistem berbasis web yang dapat digunakan oleh mahasiswa dan dosen pembimbing untuk memeriksa tingkat kemiripan judul dan abstrak skripsi secara lebih praktis dan efisien [10].

Penelitian ini memberikan kontribusi penting dalam pengembangan sistem yang membantu perguruan tinggi dalam memerangi plagiarisme, khususnya dalam konteks pengajuan judul dan abstrak skripsi. Dengan sistem yang diusulkan, mahasiswa dapat dengan cepat mengetahui apakah judul yang mereka ajukan mirip dengan judul skripsi yang sudah ada. Selain itu, dosen pembimbing dapat memantau kemiripan skripsi secara lebih efisien, memungkinkan mereka untuk memberikan bimbingan yang lebih tepat kepada mahasiswa. Sistem ini juga memberikan kontribusi terhadap literatur akademik dengan menggabungkan dua metode yang telah terbukti efektif dalam deteksi plagiarisme, yaitu WInnowing dan Dice Similarity, serta menambahkan konteks penggunaannya pada aplikasi berbasis web untuk meningkatkan kemudahan penggunaan bagi mahasiswa dan dosen di lingkungan akademik.

## 2. METODOLOGI PENELITIAN

Dalam prosedur penelitian ini menggambarkan tahapan–tahapan yang dilakukan penulis dalam melakukan penelitian ini ditunjukkan pada gambar 1 tahapan penelitian.



**Gambar 1.** Tahapan Penelitian

### 2.1 Identifikasi Masalah

Pada tahap identifikasi masalah, fokus utama dari penelitian ini adalah bagaimana merancang dan membangun sistem berbasis web yang mampu mengukur tingkat kemiripan antara judul dan abstrak skripsi. Sistem ini akan menggunakan algoritma Winnowing dan Dice Similarity sebagai metode utama dalam mendeteksi plagiarisme. Permasalahan yang dihadapi adalah bagaimana cara implementasi algoritma-algoritma tersebut secara efektif untuk mendukung pengukuran kemiripan secara akurat serta bagaimana memastikan sistem dapat diintegrasikan secara optimal dalam konteks pencegahan plagiarisme, khususnya di lingkungan akademik.

### 2.2 Pengumpulan Data

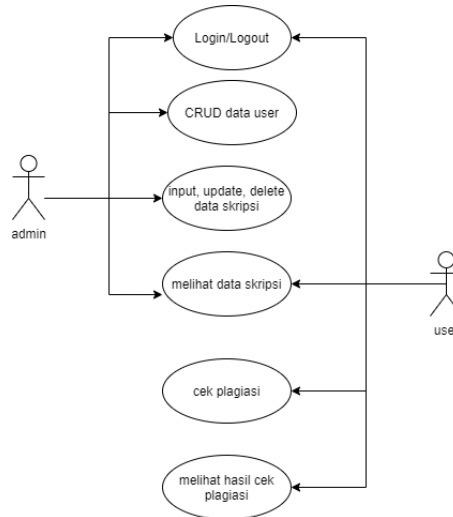
Metode pengumpulan data yang digunakan dalam penelitian ini meliputi observasi, studi literatur, dan wawancara. Observasi dilakukan dengan mengamati secara langsung sistem yang dibutuhkan di tempat penelitian dan mengumpulkan 30 data skripsi sebagai data uji atau pembanding. Studi literatur dilakukan dengan mencari informasi dan teori yang relevan dengan topik penelitian, guna mendukung pemahaman teori yang berkaitan dengan penggunaan algoritma Winnowing dan Dice Similarity. Selain itu, wawancara juga dilakukan dengan panitia skripsi dari Fakultas Teknologi Informasi Universitas Hasyim Asy'ari untuk memperoleh informasi lebih mendalam mengenai sistem pengukuran kemiripan skripsi yang sedang diteliti.

### 2.3 Analisa Kebutuhan Sistem

Kebutuhan fungsional sistem mencakup beberapa fungsi penting, yaitu untuk admin, sistem harus mendukung log in, log out, pengaturan akun, serta pengelolaan data skripsi yang meliputi create, edit, dan delete. Bagi mahasiswa, sistem harus memungkinkan log in, log out, pengaturan akun, pengecekan kemiripan judul dan abstrak, serta memberikan akses untuk melihat hasil pengecekan dan daftar skripsi yang terindikasi plagiat. Sistem juga harus dapat menampilkan data skripsi, mengupload file skripsi, melakukan pengecekan kemiripan judul dan abstrak, serta menampilkan persentase kemiripan. Sementara itu, kebutuhan non-fungsional meliputi perangkat keras dengan spesifikasi minimal prosesor Intel Core i3, RAM 2 GB, dan hard disk 320 GB. Dari sisi perangkat lunak, sistem memerlukan aplikasi seperti Visual Studio Code, Sublime Text, XAMPP, Opera Browser, dan sistem operasi Windows 7 atau 10.

### 2.4 Perancangan Sistem

Pada gambar 2 use case system menjelaskan bahwa sistem dapat di akses oleh dua user yaitu admin dan user (mahasiswa). User admin dapat diakses dengan melakukan log in terlebih dahulu, admin dapat melakukan cread, read, update, delete data mahasiswa dan skripsi. Untuk mengakses user (mahasiswa) user diharuskan log in terlebih dahulu, setelah login mahasiswa dapat melihat data skripsi, melakukan cek kemiripan judul dan abstrak, serta dapat melihat hasil cek kemiripan. metode perhitungan untuk pengukuran kemiripan adalah algoritma winnowing dimana suatu algoritme untuk melakukan proses pengecekan kesamaan kata untuk mendeteksi plagiarisme dalam dua atau lebih teks yang dibandingkan. Secara teknis algoritma winnowing merupakan ekstensi dari implementasi algoritma rabin-krap fingerprint [20].



**Gambar 2.** Use Case Sistem

Berikut adalah uraian metode yang digunakan dalam penelitian ini:

a. **Winnowing Algorithm**

Algoritma Winnowing adalah metode hashing yang sering digunakan untuk mendeteksi plagiarisme atau kemiripan teks. Algoritma ini bekerja dengan membagi teks menjadi beberapa bagian kecil yang disebut "k-grams" atau "windows." Setiap k-gram diberi nilai hash, dan nilai-nilai hash ini kemudian digunakan untuk membandingkan teks. Winnowing memilih beberapa nilai hash yang paling kecil sebagai "fingerprint" teks tersebut. Keunggulan utama dari Winnowing adalah kemampuannya untuk bekerja dengan teks yang memiliki sedikit perbedaan dalam urutannya, sehingga ideal untuk mendeteksi kemiripan yang tidak persis sama [19].

b. **Dice Similarity**

Dice Similarity: Dice Similarity Coefficient sering digunakan untuk menghitung kesamaan antara dua set dalam berbagai aplikasi termasuk teks. Pengukuran ini digunakan dalam penelitian ini untuk memberikan nilai yang mencerminkan kemiripan antara judul dan abstrak skripsi berdasarkan elemen yang sama yang ditemukan dalam teks [15].

Berikut merupakan tahapan-tahapan dari algoritma winnowing dan dice similarity

a. **Whitespace Insensitive**

Whitespace insensitive adalah salah satu tahapan dari algoritme deteksi kemiripan dimana pada tahap ini akan membuang tanda baca, spasi dan karakter-karakter yang tidak relevan (@, #, \$, %, ^, & dan lainnya), Contoh : jang an berisik !!! Akan dirubah menjadi : j ang an berisik [21]

b. **Pembentukan rangkaian n-gram**

Pada tahap ini akan dibentuk rangkaian karakter sepanjang n dari hasil proses whitespace insentive, berdasarkan teks sebelumnya akan di ubah ke deretan n semisal dengan ukuran k = 5 maka akan menjadi “ J ang a ng an ng an b gan be an ber n beri beris erisik” [22] .

c. **Perhitungan Nilai Hash pada setiap n-gram**

Pada tahapan ini akan dilakukan perhitungan nilai hash dari teks yang terbentuk dari proses sebelumnya dengan melakukan rolling hash. Rolling hash merupakan suatu cara untuk mentransformasikan sebuah string menjadi suatu nilai yang unik dengan panjang tertentu dimana berfungsi sebagai penanda tersebut. Fungsi untuk mendapatkan nilai disebut fungsi hash sedangkan nilai yang dihasilkan disebut nilai hash. Berikut fungsi hash [23].

$$H_{Ck} = C_1 * b^{(k-1)} + C_2 * b^{(k-2)} + ... + C_k * b^{(k-k)} \tag{1}$$

Untuk melakukan enkripsi sebuah pesan, karakter-karakter dalam pesan tersebut diubah menjadi nilai ASCII yang direpresentasikan sebagai variabel C. Basis yang digunakan untuk enkripsi adalah b, yang merupakan sebuah bilangan prima. Selain itu, k merepresentasikan banyaknya karakter dalam pesan tersebut. Dengan memanfaatkan nilai ASCII dari masing-masing karakter dan basis prima, pesan dapat diubah menjadi bentuk terenkripsi, di mana setiap karakter dalam pesan dikonversi berdasarkan formula yang menggabungkan nilai C, b, dan k. Proses ini menghasilkan output terenkripsi yang sulit ditebak tanpa mengetahui basis prima dan karakter asli pesan tersebut.

d. **Pembentukan window dari nilai hash**

e. **Pemilihan nilai terkecil dari setiap window untuk dijadikan fingerprint**

f. **Metode Dice coefficient memiliki keunggulan dimana proses komputasinya yang tidak terlalu berat tetapi memiliki hasil yang akurat [24]. Berikut perumusan metode dice similarity :**

$$S(a,b) = \frac{2|a \cap b|}{|a| + |b|} \tag{2}$$



Kemiripan antara dua teks, yang dinotasikan sebagai S(a,b), dihitung dengan membandingkan fingerprint dari kedua teks tersebut. Nilai |a| merepresentasikan jumlah fingerprint pada teks pertama, sedangkan |b| merepresentasikan jumlah fingerprint pada teks kedua. Sementara itu, |a∩b| merupakan jumlah fingerprint yang sama antara kedua teks. Semakin banyak fingerprint yang sama, semakin tinggi nilai kemiripan (S(a,b)) antara kedua teks, yang menggambarkan seberapa mirip kedua teks tersebut berdasarkan fingerprint mereka.

g. Penentuan Nilai Kemiripan

Mengacu pada penelitian yang mengidentifikasi lima karakteristik penilaian persentase similarity [25], hasil kemiripan dokumen dapat diinterpretasikan sebagai berikut: nilai 0% menunjukkan bahwa kedua dokumen tersebut benar-benar berbeda, nilai di bawah 15% berarti kedua dokumen teridentifikasi memiliki sedikit kesamaan, persentase antara 15-50% menandakan bahwa dokumen tersebut teridentifikasi memiliki plagiarisme tingkat sedang, jika persentase lebih dari 50%, ini menandakan bahwa dokumen mendekati plagiarisme, dan nilai 100% menunjukkan plagiarisme penuh, di mana keseluruhan isi kedua dokumen sama.

2.5 Implementasi

Implementasi metode menggunakan framework Laravel PHP bertujuan mempercepat pengembangan dan meningkatkan keamanan serta efisiensi sistem. Algoritma WInnowing dan Dice Similarity akan diterapkan untuk mengukur kemiripan judul dan abstrak skripsi, yang diintegrasikan dalam struktur Laravel. Fitur seperti routing, middleware, dan Eloquent ORM memudahkan pengelolaan database, sehingga admin dapat mengelola data skripsi dengan lebih efisien melalui antarmuka yang user-friendly.

2.6 Pengujian

Menguji sistem dengan metode Blackbox untuk memastikan fungsionalitas sesuai.

3. HASIL DAN PEMBAHASAN

3.1 Hasil penelitian

Hasil penelitian sistem pengukuran tingkat kemiripan judul dan abstrak skripsi berbasis web menggunakan algoritma WInnowing dan Dice Similarity menunjukkan beberapa fitur utama. Pertama, admin dapat mengelola data mahasiswa dan skripsi, termasuk menambah, mengubah, dan menghapus data tersebut. Kedua, mahasiswa dapat mengubah data diri, melihat data skripsi mahasiswa sebelumnya yang sudah dinyatakan lulus, serta melakukan pengecekan plagiasi pada judul dan abstrak skripsi. Hasil pengecekan ditampilkan dalam bentuk daftar beserta persentase kemiripan dari setiap data skripsi yang ada. Tabel 2 berikut adalah contoh data judul dan abstrak.

Tabel 1. Data Judul Dan Abstrak

Kategori	Judul	Abstrak
Judul dan Abstrak Uji	Rancang Bangun Sistem Monitoring Laporan Quality Control Metode Decision Tree	Pada kasus permasalahan di PT. Sumber Graha Sejahtera adalah perusahaan yang bergerak di bidang industri kayu lapis pada permasalahan ini PT. Sumber Graha Sejahtera sumber graha sejahtera sedang mengalami krisis anggaran untuk pembelian perlengkapan terutamanya devisi Quality Control yang pendataan sangat penting karena mendata kualitas barang di lapangan.
Judul dan Abstrak Pemanding	Rancang Bangun Sistem Informasi Objek Wisata Dan Hotel Kabupaten Jombang	Wisata merupakan seluruh kegiatan yang berhubungan dengan perjalanan wisata, sebelum maupun selama perjalanan dan kembali ke tempat asal. Setiap daerah memiliki objek wisata, begitu pula dengan Jombang yang dikenal sebagai Kota Santri Jombang juga memiliki wisata yang sangat berpotensi, dimana terdapat banyak lokasi wisata dan hotel terdekat.

Dalam tahap ini dilakukan pengujian perhitungan dari algoritma winnowing dan dice similarity untuk menentukan seberapa besar plagiasi dari judul dan abstrak

a. Hasil judul uji dan judul pemanding

1. Jumlah Fingerprints judul uji = 60
2. Jumlah Fingerprints judul pemanding = 55
3. Union (Gabungan) Fingerprints judul uji dan judul pemanding = 115
4. Intersection (fingerprints yang sama) = 13
5. Persentase Plagiarisme Koefisien Dice = (Intersection \* 2 / Union) \* 100, Jadi (13 \* 2/115) \* 100 = 22.608 %

6. Hasil Plagiarisme = 22.608 % dilakukan pembulatan menjadi 22.61%
- b. Hasil abstrak uji dan abstrak pembandingan
  1. Jumlah Fingerprints abstrak uji = 656
  2. Jumlah Fingerprints abstrak pembandingan = 1285
  3. Union (Gabungan) Fingerprints abstrak uji dan abstrak pembandingan = 1941
  4. Intersection (fingerprints yang sama) = 455
  5. Persentase Plagiarisme Koefisien Dice =  $(\text{Intersection} * 2 / \text{Union}) * 100$ , Jadi  $(455 * 2 / 1941) * 100 = 46.883 \%$
  6. Hasil Plagiarisme = 46.883 % dilakukan pembulatan menjadi 46.88%
- c. Akumulasi hasil plagiarisme  
Setelah didapatkan hasil dari perbandingan judul uji dan judul pembandingan, abstrak uji dan abstrak pembandingan, kemudian hasil tersebut akan di akumulasikan dengan di ambil rata-ratanya seperti berikut :

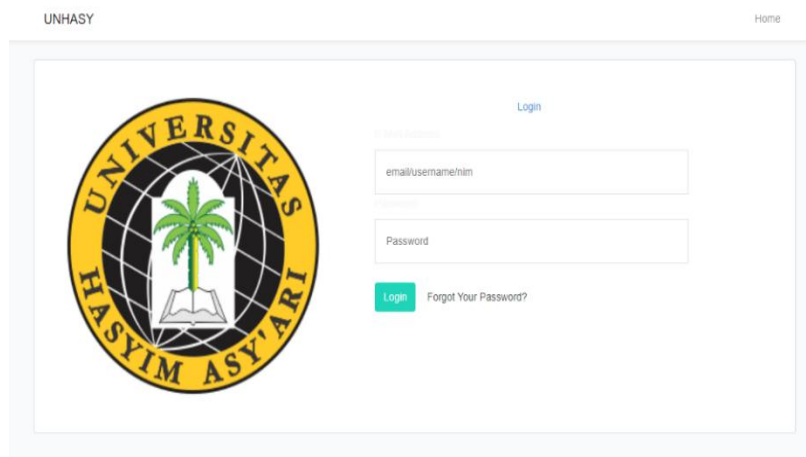
$$\text{Rumus rata-rata: } m = \frac{\text{jumlah data}}{\text{banyak data}} \quad (3)$$

$$\text{Maka } m = \frac{22.61 + 46.88}{2} = 34.745\%$$

Jadi dapat disimpulkan bahwa cek plagiasi antara judul dan abstrak uji dengan judul dan abstrak data Pembandingan bernilai 34.745% dapat dikategorikan dalam plagiat tingkat sedang.

### 3.1.1 Halaman Login

Gambar 7 menampilkan halaman login dari sistem yang digunakan oleh Universitas Hasyim Asyari (UNHASY).

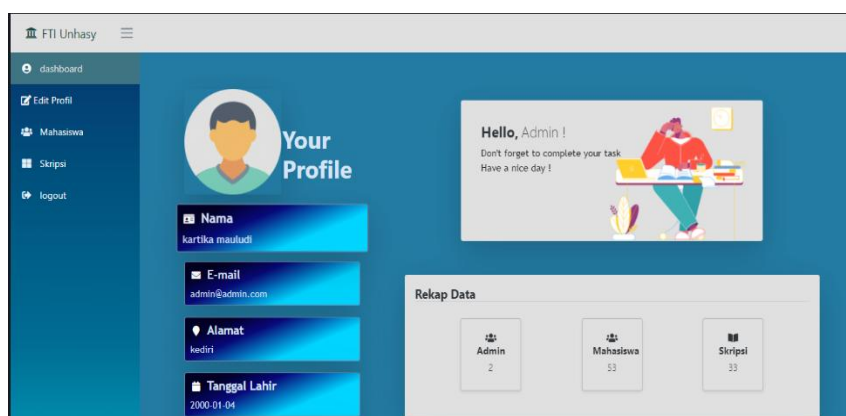


**Gambar 3.** Halaman Login

Pada Gambar 7 pengguna diminta untuk memasukkan username dan password mereka untuk mengakses sistem. Selain itu, terdapat opsi untuk pengguna yang lupa password, yang memungkinkan mereka untuk memulihkan akses ke akun mereka.

### 3.1.2 Halaman Admin

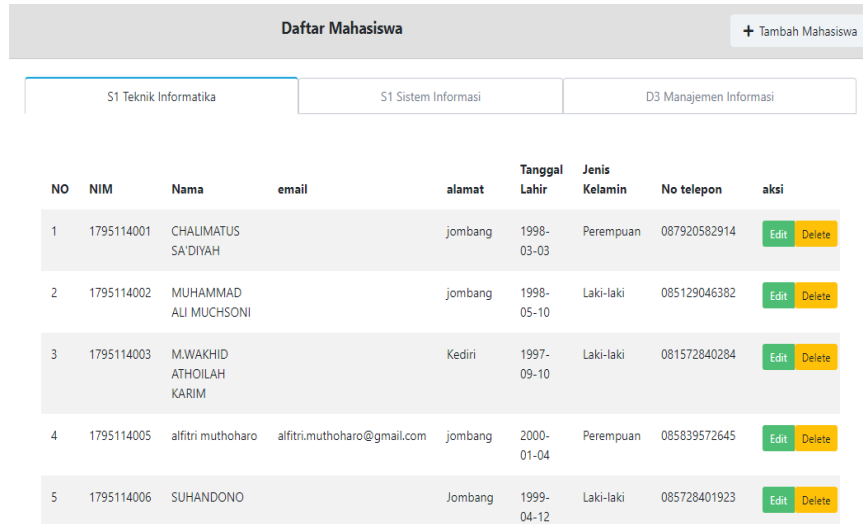
Gambar 8 adalah halaman dashboard admin dapat melihat informasi mengenai profil pribadi, data jumlah admin, data jumlah mahasiswa, dan data jumlah skripsi yang ada pada sistem.



**Gambar 4.** Halaman Admin

### 3.1.3 Halaman Data User

Gambar 9 merupakan halaman daftar mahasiswa berisi data mahasiswa yang tercatat pada sistem, data berisi NIM, nama, email(optional), tanggal lahir, jenis kelamin, no telephone dan data dibedakan menurut program studi yang diambil.



NO	NIM	Nama	email	alamat	Tanggal Lahir	Jenis Kelamin	No telepon	aksi
1	1795114001	CHALIMATUS SA'DIYAH		Jombang	1998-03-03	Perempuan	087920582914	Edit Delete
2	1795114002	MUHAMMAD ALI MUCHSONI		Jombang	1998-05-10	Laki-laki	085129046382	Edit Delete
3	1795114003	M.WAKHID ATHOILAH KARIM		Kediri	1997-09-10	Laki-laki	081572840284	Edit Delete
4	1795114005	alfitri muthoharo	alfitri.muthoharo@gmail.com	Jombang	2000-01-04	Perempuan	085839572645	Edit Delete
5	1795114006	SUHANDONO		Jombang	1999-04-12	Laki-laki	085728401923	Edit Delete

Gambar 5. Halaman Data User

### 3.1.4 Halaman Data Skripsi

Gambar 10 merupakan halaman yang berisi deretan data skripsi dari mahasiswa mahasiswa yang sudah menyelesaikan studi S1 di fakultas teknologi informasi dimana di dalamnya berisi nama, Nim, nama pembimbing, abstrak skripsi dan jurnal skripsi.



NO	NIM	nama penulis	Judul	Prodi	Aksi
1	1495124003	Johan Fikandda	SISTEM PENDUKUNG KEPUTUSAN CALON PENERIMA BEASISWA KURANG MAMPU PADA SMK SULTAN AGUNG 1 MENGGUNAKAN METODE SIMPLE ADDITIVE WEIGHTING	S1 Sistem Informasi	Detail Edit Delete
2	1595114009	Nunggi Libasut Taqwa	SISTEM PREDIKSI PRODUKSI PADI DI PROVINSI JAWA TIMUR MENGGUNAKAN EXPONENTIAL SMOOTHING BERBASIS WEB	S1 Teknik Informatika	Detail Edit Delete
3	1595114019	Bayu Putra	RANCAANG BANGUN APLIKASI KOREKSI LEMBAR JAWABAN KOMPUTER MENGGUNAKAN METODE DETEKSI TEPI CANNY	S1 Teknik Informatika	Detail Edit Delete

Gambar 6. Halaman Data Skripsi

### 3.1.5 Halaman Mahasiswa

Gambar 11 merupakan ketika mahasiswa berhasil login sistem halaman pertamanya yang muncul merupakan halaman dashboard mahasiswa berisi informasi singkat mengenai sistem.



Gambar 7. Halaman Mahasiswa

### 3.1.6 Halaman Cek Plagiasi

Gambar 12 merupakan halaman masukan dari judul dan abstrak yang nantinya judul dan abstrak yang telah di isi akan diproses oleh sistem dengan algoritma winnowing dan dice similarity.

**Gambar 8.** Halaman Cek Plagiasi

### 3.1.7 Halaman Hasil Cek Plagiasi

Gambar 13 merupakan halaman hasil cek plagiasi, pada halaman ini mahasiswa dapat melihat hasil cek plagiasi dengan menampilkan plagiasi tertinggi dan akan diperlihatkan judul dan abstrak pembanding.

#	Source	Comparator
Judul	Rancang Bangun Sistem Monitoring Laporan Quality Control Metode Decision Tree	RANCANG BANGUN APLIKASI KOREKSI LEMBAR JAWABAN KOMPUTER MENGGUNAKAN METODE DETEKSI TEPI CANNY
Abstrak	Pada kasus permasalahan di PT. Sumber Graha Sejahtera adalah perusahaan yang bergerak di bidang industri kayu lapis pada permasalahan ini PT. Sumber Graha Sejahtera sumber graha sejahtera sedang mengalami krisis anggaran untuk pembelian perlengkapan utamanya devisi Quality Control yang pendataan sangat penting karena mendata kualitas barang di lapangan. Sering terjadi kehilangan data dan kekurangan data di karenakan pendataan menggunakan sistem manual kertas dan sering terjadi kekurangan kertas mengakibatkan pendataan tidak berjalan secara terus-menerus di karenakan kekurangan kertas maka dengan ini di buatlah Rancang bangun system Monitoring Quality Control Barang berbasis web agar dapat mempermudah pendataan dan dapat menghindari data pemecakan yang kurang	Penilaian Akhir Semester adalah salah satu bagian dari proses evaluasi hasil belajar yang menggunakan Lembar Jawaban Komputer (LJK) sebagai tempat siswa menuliskan jawaban. Umumnya LJK akreditasi menggunakan scanner tertentu yang harganya relatif mahal, sehingga beberapa sekolah memilih untuk mengoreksinya secara manual termasuk di MA Ar-Rahman Sumojono. Untuk mengatasi masalah yang ada, maka penelitian ini adalah salah satu solusi yang memanfaatkan smartphone dan teknologi pengolahan citra digital khususnya menggunakan metode deteksi tepi canny. Rancang bangun aplikasi ini nantinya akan mengkombinasikan metode deteksi tepi canny untuk mencari tepi dari citra dan fungsi hough circle transformation untuk mendeteksi lingkaran yang telah dihitamkan. Rancang bangun aplikasi ini juga menggunakan nilai parameter canny yang dapat dirubah oleh user, mengingat tingkat kecerahan pada setiap pengambilan gambar berbeda-beda. Hasil dari citra yang telah terdeteksi jawabannya akan dilakukan proses tracking untuk menentukan jawaban yang dipilih. Jawaban dari siswa nantinya akan dicocokkan dengan kunci jawaban yang telah disimpan dalam database SQLite. Hasil dari pengujian terhadap aplikasi yang telah dibuat menggunakan 30 sampel LJK, didapatkan rata-rata persentase akurasi kesesuaian hasil deteksi jawaban dengan jawaban sesungguhnya di LJK adalah 91,41 %. &nbsp;tsnp;

**Gambar 9.** Halaman Hasil Cek Plagiasi

## 3.2 Evaluasi

Pada hasil pengujian tabel 1 perhitungan manual dan pengujian pada sistem menggunakan data uji dan data pembanding menggunakan data masukan, n-gram = 5 , window = 5 dan bilangan prima = 2 yang sama. Hasil perhitungan manual dan pengujian sistem menunjukkan hasil yang sama. Hasil dapat dilihat pada pada tabel 1 dan tabel 2 di bawah ini:

**Tabel 1.** Hasil Perhitungan Manual

No	Data uji	Data pembanding	N-Gram	Window	B.Prima	Hasil
1	Judul dan abstrak masukan	Judul dan abstrak id_8	5	5	2	<b>34,475%</b>

**Tabel 2.** Hasil Pengujian Sistem

No	Id_Data	N-Gram	Window	B.Prima	Hasil
1	Id_20	5	5	2	27.715%
2	Id_21	5	5	2	25.255%
3	Id_22	5	5	2	32.93%
4	Id_23	5	5	2	27.845%
5	<b>Id_24</b>	<b>5</b>	<b>5</b>	<b>2</b>	<b>34.745%</b>
6	Id_25	5	5	2	32.905%
7	Id_26	5	5	2	32.09%
8	Id_27	5	5	2	27.56%

Pada algoritma winnowing tidak ada ketentuan khusus mengenai berapakah nilai gram, window dan prima yang harus digunakan, jadi disini penulis melakukan pengujian untuk mengetahui nilai berapakah yang paling efisien. Hasil pengujian dapat dilihat pada tabel 3, tabel 4, tabel 5 dibawah ini.

**Tabel 3.** Pengujian dengan nilai gram = 4, window = 4, prima = 3

No	ID-Data	Hasil
1	Id_1	27.215%
2	Id_2	24.18%
3	Id_3	21.59%
4	Id_4	26.875%
5	Id_5	37.61%
6	Id_6	28.43%
7	Id_7	26.33%
8	Id_8	29.465%
9	Id_9	22.86%
10	Id_10	23.205%

Tabel 3 memaparkan hasil pengujian dengan nilai gram = 4, window = 4, dan prima = 3, dari pengujian di atas didapat nilai tertinggi pada id\_5 = 37,61% dan nilai terendah pada id\_3 = 21,59% .

**Tabel 4.** Pengujian dengan nilai gram = 4, window = 3, prima =3

No	Id_Data	Hasil
1	Id_1	25.635%
2	Id_2	25.37%
3	Id_3	21.65%
4	Id_4	27.865%
5	Id_5	38.16%
6	Id_6	29.115%
7	Id_7	27.5%
8	Id_8	28.665%
9	Id_9	23.725%
10	Id_10	23.305%

Tabel 4. memaparkan hasil pengujian dengan nilai gram = 4, window = 3, dan prima = 3, dari pengujian di atas didapat nilai tertinggi pada id\_5 = 38.16% dan nilai terendah pada id\_3 = 21.65%

**Tabel 5.** Pengujian dengan nilai gram = 5, window = 4, prima = 3

No	Id_Data	Hasil
1	Id_1	10.82%
2	Id_2	17.46%
3	Id_3	15.21%
4	Id_4	16%
5	Id_5	17.665%
6	Id_6	17.885%
7	Id_7	15.57%
8	Id_8	15.225%
9	Id_9	14.75%
10	Id_10	16.87%

Tabel 5 Pengujian dengan nilai gram = 5, window = 4, prima = 3, memaparkan hasil pengujian dengan nilai gram = 4, window = 3, dan prima = 3, dari pengujian di atas didapat nilai tertinggi pada id\_6 = 17.885% dan nilai terendah pada id\_1 = 10.82%, Berdasarkan hasil pengujian maka diperoleh tiga hasil nilai, dari hasil tersebut terdapat nilai terkecil yang diperoleh dari nilai masukan gram = 5, window = 4 dan prima = 3 yaitu 10.82%, sedangkan hasil terbesar yaitu 38.16% diperoleh dari nilai masukan gram = 4, window = 3 dan prima = 3, dari kedua hasil ini dapat diambil kesimpulan bahwa semakin besar nilai dari ngram, window dan prima maka hasil yang diperoleh akan semakin kecil begitu juga sebaliknya semakin kecil nilai gram, window dan prima maka hasil yang didapat akan semakin tinggi.

## 4. KESIMPULAN

Berdasarkan hasil penelitian, penulis menyimpulkan bahwa sistem pengukuran kemiripan judul dan abstrak skripsi berbasis web berhasil dikembangkan menggunakan algoritma winnowing dan dice similarity. Sistem ini dirancang



dengan konsep OOP (Object Oriented Programming) dan memanfaatkan PHP serta MySQL sebagai database. Meskipun tidak ada standar baku untuk nilai gram, window, dan prima, algoritma winnowing diterapkan dengan sukses, di mana pengujian dilakukan untuk menemukan kombinasi optimal yang menghasilkan tingkat kemiripan antara 10.82% hingga 38.16%. Dibandingkan dengan algoritma Jaccard Coefficient yang digunakan dalam penelitian sebelumnya, dice similarity menunjukkan peningkatan kemiripan sebesar 13%. Untuk pengembangan lebih lanjut, disarankan untuk mencoba algoritma lain guna membandingkan akurasi dalam pengecekan plagiarisme, menerapkan algoritma winnowing pada objek penelitian yang lebih kecil, serta menambahkan fitur trace back untuk mendeteksi kata yang sama dan menerapkan stemming dalam bahasa Indonesia dan Inggris.

## REFERENCES

- [1] N. D. P. Wijaya, D. I. Tsamroh, J. W. Dika, C. Yazirin, D. A. Ross, and A. D. Kristanto, 'Sosialisasi Dan Pelatihan Penggunaan Mendeley Untuk Persiapan Mahasiswa Dalam Penulisan Skripsi', *Vivabio J. Pengabd. Multidisiplin*, vol. 4, no. 1, p. 12, 2022, doi: 10.35799/vivabio.v4i1.40876.
- [2] M. Melisa, S. T. Maulia, and F. Nugraha, 'Mendeley Sebagai Upaya Menghindari Plagiarisme Dan Peningkatan Kualitas Penulisan Karya Ilmiah Mahasiswa', *Edu Sos. J. Pendidik. Ilmu Pengetah. Sos.*, vol. 3, no. 1, pp. 77–82, 2023, doi: 10.22437/jeso.v3i1.24646.
- [3] C. D. Syafriani and S. Fachrurrazi, 'Sistem Informasi Pengajuan Judul Tugas Akhir Dan Konsultasi Skripsi Berbasis Web Di Teknik Kimia Universitas Malikussaleh Lhokseumawe', *Sisfo J. Ilm. Sist. Inf.*, vol. 4, no. 1, 2020, doi: 10.29103/sisfo.v4i1.6274.
- [4] A. Kurniawan, 'Pembuatan Sistem Administrasi Tugas Akhir Berbasis Teknologi Informasi Pada Program Studi S1 Teknik Kimia UPN "Veteran" Yogyakarta', *Telematika*, vol. 17, no. 1, p. 38, 2020, doi: 10.31315/telematika.v17i1.3391.
- [5] A. Syam, A. Asniati, and W. T. W. Tita, 'Penerapan Sistem Informasi Pada Layanan Pengajuan Dan Monitoring Judul Tugas Akhir Mahasiswa', *J. Inform.*, vol. 10, no. 1, p. 56, 2021, doi: 10.55340/jiu.v10i1.567.
- [6] D. Darmanto, 'Sistem Deteksi Plagiarisme Tugas Akhir Mahasiswa Berbasis Natural Language Processing Menggunakan Algoritma Jaro-Winkler Dan TF-IDF', *Smart Comp Jurnalnya Orang Pint. Komput.*, vol. 13, no. 1, 2024, doi: 10.30591/smartcomp.v13i1.6375.
- [7] D. G. Baskoro, 'Dampak Pelatihan Literasi Informasi Pada Persepsi Mahasiswa Terhadap Plagiarisme', *Libr. J. Perpust.*, vol. 10, no. 2, p. 205, 2022, doi: 10.21043/libraria.v10i2.16725.
- [8] I. Idris, A. Subekti, and W. D. Sulisty, 'Pelatihan Karya Tulis Ilmiah Dan Pengelolaan Jurnal Elektronik Sebagai Media Halaqoh Ilmiah Di Pesantren Luhur', *J. Praksis Dan Dedik. Sos. Jpds*, vol. 3, no. 1, p. 11, 2020, doi: 10.17977/um032v3i1p11-20.
- [9] D. Armiady, 'Pengembangan Sistem Informasi Pengajuan Judul Skripsi Menggunakan Algoritma Winnowing', *J. Tika*, vol. 8, no. 2, pp. 117–123, 2023, doi: 10.51179/tika.v8i2.2066.
- [10] M. Mentari, I. F. Rozi, and M. P. Rahayu, 'Cross-Language Text Document Plagiarism Detection System Using Winnowing Method', *J. Appl. Intell. Syst.*, vol. 7, no. 1, pp. 44–57, May 2022, doi: 10.33633/jais.v7i1.5950.
- [11] A. Cahyapratama, K. R. Sungkono, and R. Sarno, 'Gap analysis business process model by using structural similarity', *Indones. J. Electr. Eng. Comput. Sci.*, vol. 18, no. 1, p. 124, Apr. 2020, doi: 10.11591/ijeecs.v18.i1.pp124-134.
- [12] E. H. A. Prastyo and M. Faisal, 'Meningkatkan Akurasi Deteksi Berita Palsu dengan Pendekatan Berbasis Lexicon dan LSTM melalui Text Preprocessing dan Model Training', *J. MEDIA Inform. BUDIDARMA*, vol. 8, no. 3, p. 1746, Jul. 2024, doi: 10.30865/mib.v8i3.7847.
- [13] S. Sakira Kamaruddin, Y. Yusof, N. Azzah Abu Bakar, M. Ahmed Tayie, and G. Abdulsattar A. Jabbar Alkubaisi, 'Graph-based Representation for Sentence Similarity Measure : A Comparative Analysis', *Int. J. Eng. Technol.*, vol. 7, no. 2.14, p. 32, Apr. 2018, doi: 10.14419/ijet.v7i2.14.11149.
- [14] E. Prastyo, 'Implementation of Web Scraping on News Sites Using the Supervised Learning Method', *İlköğretim Online*, vol. 20, no. 3, Jan. 2021, doi: 10.17051/ilkonline.2021.03.43.
- [15] M. Yeung, L. Rundo, Y. Nan, E. Sala, C.-B. Schönlieb, and G. Yang, 'Calibrating the Dice Loss to Handle Neural Network Overconfidence for Biomedical Image Segmentation', *J. Digit. Imaging*, vol. 36, no. 2, pp. 739–752, Dec. 2022, doi: 10.1007/s10278-022-00735-3.
- [16] S. Sunardi, A. Yudhana, and I. A. Mukaromah, 'IMPLEMENTASI DETEKSI PLAGIARISME MENGGUNAKAN METODE N-GRAM DAN JACCARD SIMILARITY TERHADAP ALGORITMA WINNOWING', *Transmisi*, vol. 20, no. 3, p. 105, Oct. 2018, doi: 10.14710/transmisi.20.3.105-110.
- [17] M. Farouk, 'Measuring text similarity based on structure and word embedding', *Cogn. Syst. Res.*, vol. 63, pp. 1–10, Oct. 2020, doi: 10.1016/j.cogsys.2020.04.002.
- [18] N. Alamsyah, 'PERBANDINGAN ALGORITMA WINNOWING DENGAN ALGORITMA RABIN KARP UNTUK MENDETEKSI PLAGIARISME PADA KEMIRIPAN TEKS JUDUL SKRIPSI', *Technol. J. Ilm.*, vol. 8, no. 3, p. 124, Jul. 2017, doi: 10.31602/tji.v8i3.1116.
- [19] I. Widaningrum, D. Mustikasari, R. Arifin, and H. A. Pratiwi, 'Evaluation of the accuracy of winnowing, rabin karp and knuth morris pratt algorithms in plagiarism detection applications', *J. Phys. Conf. Ser.*, vol. 1517, no. 1, p. 012093, Apr. 2020, doi: 10.1088/1742-6596/1517/1/012093.
- [20] D. Leman, M. Rahman, F. Ikorasaki, B. S. Riza, and M. B. Akbbar, 'Rabin Karp And Winnowing Algorithm For Statistics Of Text Document Plagiarism Detection', in *2019 7th International Conference on Cyber and IT Service Management (CITSM)*, Jakarta, Indonesia: IEEE, Nov. 2019, pp. 1–5. doi: 10.1109/CITSM47753.2019.8965422.
- [21] Y. Wan, F. Ren, L. Yao, and Y. Zhang, 'Research on Scene Chinese Character Recognition Method Based on Similar Chinese Characters', in *2020 2nd International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI)*, Taiyuan, China: IEEE, Oct. 2020, pp. 459–463. doi: 10.1109/MLBDBI51377.2020.00097.



- [22] W. Qi et al., ‘ProphetNet: Predicting Future N-gram for Sequence-to-SequencePre-training’, in Findings of the Association for Computational Linguistics: EMNLP 2020, Online: Association for Computational Linguistics, 2020, pp. 2401–2410. doi: 10.18653/v1/2020.findings-emnlp.217.
- [23] A. K. Saputra, E. Erlangga, and T. Tanjung, ‘Analisis Kinerja Algoritma Winnowing pada Pendeteksian Plagiarisme’, *Explore J. Sist. Inf. Dan Telematika*, vol. 13, no. 2, p. 125, Dec. 2022, doi: 10.36448/jsit.v13i2.2845.
- [24] M. Wang and D. Li, ‘An Automatic Segmentation Method for Lung Tumor Based on Improved Region Growing Algorithm’, *Diagnostics*, vol. 12, no. 12, p. 2971, Nov. 2022, doi: 10.3390/diagnostics12122971.
- [25] M. Nurzaman and I. D. Nursasih, ‘Pengaruh Efektivitas Penilaian Sistem Digital Pertandingan Pencak Silat Berbasis Komputer Dengan Sistem Penilaian Manual’, *J. Keolahragaan*, vol. 7, no. 1, p. 48, Apr. 2021, doi: 10.25157/jkor.v7i1.5314.