



**UNIVERSITAS INDO GLOBAL MANDIRI**

**KOMPARASI *JACCARD SIMILARITY* DAN *COSINE  
SIMILARITY* DALAM MENDETEKSI PLAGIARISME  
ABSTRAK JURNAL**

**SKRIPSI**

**Diajukan Sebagai Syarat Untuk Menyelesaikan  
Pendidikan Program Strata - 1  
Pada Program Studi Teknik Informatika**

**Oleh :**

**Devi Yusrina**

**2019.11.0040**

**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS INDO GLOBAL MANDIRI  
2023**

**KOMPARASI *JACCARD SIMILARITY* DAN *COSINE SIMILARITY* DALAM MENDETEKSI PLAGIARISME  
ABSTRAK JURNAL**



**SKRIPSI**

**Diajukan Sebagai Syarat Untuk Menyelesaikan  
Pendidikan Program Strata – 1  
Pada Program Studi Teknik Informatika**

**Oleh :**

**Devi Yusrina  
2019.11.0040**

**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS INDO GLOBAL MANDIRI  
2023**

# LEMBAR PENGESAHAN SKRIPSI

## LEMBAR PENGESAHAN SKRIPSI

**Komparasi *Jaccard Similarity* dan  
*Cosine Similarity* dalam Mendeteksi  
Plagiarisme Abstrak Jurnal**

Oleh :

**Devi Yusrina  
NPM : 2019.11.0040**

**Palembang , 21 Juli 2023**

**Pembimbing I**



**Rudi Heriansyah, S.T., M.Eng., P.hD.  
NIK:2022.01.0315**

**Pembimbing II**



**Lastri Widya Astuti, M.Kom  
NIK : 2003.01.0063**

Mengetahui,

**Dekan Fakultas Ilmu Komputer**



**Rudi Heriansyah, S.T., M.Eng., P.hD.  
NIK:2022.01.0315**

## LEMBAR PERSETUJUAN DEWAN PENGUJI

### LEMBAR PERSETUJUAN DEWAN PENGUJI

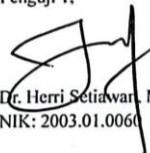
Pada hari Senin tanggal 10 Juli 2023 telah dilaksanakan ujian sidang skripsi :

Nama : Devi Yusrina  
NPM : 2019.11.0040  
Judul : *Komparasi Metode Jaccard Similarity dan Cosine Similarity*  
dalam Mendeteksi Plagiarisme Abstrak Jurnal

Oleh Prodi Teknik Informatika Fakultas Ilmu Komputer Universitas Indo Global Mandiri  
Palembang

Palembang, 21 Juli 2023

Penguji 1,



Dr. Herri Setiawan, M.Kom  
NIK: 2003.01.0060

Penguji 2,



Dewi Sartika, M.Kom.  
NIK: 2013.01.0015

Penguji 3,



Lastri Widya Astuti, M.Kom  
NIK : 2003.01.0063

Menyetujui,  
Ka. Prodi Teknik Informatika



Dewi Sartika, M.Kom  
NIK: 2013.01.0015

## SURAT KETERANGAN REVISI



SURAT KETERANGAN REVISI SKRIPSI  
PROGRAM STUDI TEKNIK INFORMATIKA (SI)  
FASILKOM UNIVERSITAS INDO GLOBAL MANDIRI

Kami yang bertanda tangan dibawah ini, menerangkan bahwa :

Nama : Devi Yusrina

NPM : 2010.11.0040

Judul : Komparasi Metode *Jaccard Similarity* dan *Cosine Similarity* dalam Mendeteksi Plagiarisme Abstrak Jurnal

Mahasiswa yang namanya tercantum diatas, telah selesai merevisi penulisan SKRIPSI

Palembang, 21 Juli 2023

Penguji 1.

Dr. Herri Setawan, M.Kom  
NIK: 2003.01.0060

Penguji 2.

Dewi Sartika, M.Kom  
NIK: 2013.01.0015

Penguji 3.

Lastri Widya Astuti, M.Kom  
NIK : 2003.01.0063

Menyetujui,  
Ka. Prodi Teknik Informatika

Dewi Sartika, M.Kom  
NIK: 2013.01.0015

**KOMPARASI *JACCARD SIMILARITY* DAN *COSINE  
SIMILARITY* DALAM MENDETEKSI PLAGIARISME  
ABSTRAK JURNAL**

**ABSTRAK**

Plagiarisme merupakan tindakan yang dilakukan dengan cara meniru suatu karya seseorang secara sengaja tanpa menyebutkan sumbernya. Plagiarisme dapat berdampak secara luas seperti merusak citra dan kredibilitas sang penulis, hingga dapat dianggap sebagai kejahatan pelanggaran hak cipta. Berbagai cara yang telah dilakukan oleh banyak orang dalam menyelesaikan masalah plagiarisme karya ilmiah seperti pembuatan aturan yang memberikan hukuman berat kepada pelaku, dan membuat suatu metode atau cara untuk mendeteksi plagiarisme karya ilmiah secara tepat dan akurat. Metode ini dilakukan dengan menggunakan sebuah program atau aplikasi *plagiarism checker* yang bekerja dengan menerapkan berbagai algoritma dalam prosesnya. Terdapat berbagai algoritma yang dapat digunakan tetapi algoritma *Jaccard Similarity* dan *Cosine Similarity* merupakan 2 (dua) algoritma yang paling populer dalam pembuatan program *plagiarisme checker*. Penelitian ini bertujuan untuk membandingkan algoritma *Jaccard Similarity* dan *Cosine Similarity* dalam melakukan proses pendeteksian plagiarisme pada sebuah karya ilmiah yang sering kali terjadi plagiarisme yaitu jurnal ilmiah. Penelitian yang dilakukan mendapatkan hasil bahwa algoritma *Jaccard Similarity* lebih baik dengan tingkat akurasi, presisi, dan *recall* yang mencapai 96% (Sembilan puluh enam persen), dibandingkan dengan algoritma *Cosine Similarity* yang hanya menunjukkan hasil akurasi dan presisi 76% (tujuh puluh enam persen).

**Kata kunci :** *Cosine Similarity*, *Jaccard Similarity*, Jurnal, plagiarisme.

**COMPARISON OF JACCARD SIMILARITY AND COSINE  
SIMILARITY IN DETECTING ABSTRACT JOURNAL  
PLAGIATION**

**ABSTRACT**

*Plagiarism is an act carried out by intentionally imitating someone's work without acknowledging its source. Plagiarism can have wide-ranging consequences, such as damaging the author's reputation and credibility, and can be considered a copyright infringement crime. Various methods have been employed by many individuals to address plagiarism issues in academic works, such as implementing strict penalties for offenders and developing accurate and precise methods or approaches to detect plagiarism. One such method involves the use of a program or application called a plagiarism checker, which employs various algorithms in its process. While there are several algorithms available, Jaccard Similarity and Cosine Similarity are two of the most popular algorithms used in the development of plagiarism checker programs. This study aims to compare the Jaccard Similarity and Cosine Similarity algorithms in detecting plagiarism in a frequently targeted form of academic work, namely scientific journals. The research findings indicate that the Jaccard Similarity algorithm outperforms the Cosine Similarity algorithm, achieving a 96% accuracy, precision, and recall rate, compared to the Cosine Similarity algorithm, which only demonstrates an accuracy and precision rate of 76%.*

**Keywords :** , Cosine Similarity, Jaccard Similarity, Journal, plagiarism,

## KATA PENGANTAR

Puji dan syukur penulis panjatkan kehadirat Allah Subhanahu Wata'ala yang telah melimpahkan rahmat dan hidayahnya, sehingga penulis dapat menyelesaikan penulisan skripsi ini dengan baik, tidak lupa shalawat serta salam selalu dilimpahkan kepada junjungan kita Nabi besar Nabi Muhammad SAW beserta keluarga sahabat para pengikut dan insyaallah kita semua hingga akhir zaman.

Skripsi yang penulis buat dengan judul “Komparasi *Jaccard Similarity* dan *Cosine Similarity* dalam Mendeteksi Plagiarisme Abstrak Jurnal” disusun guna memenuhi syarat kelulusan dalam memperoleh gelar Sarjana (S1) pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Indo Global Mandiri (UIGM) Palembang.

Penulis menyadari bahwa penyusunan skripsi ini tidak akan terwujud tanpa adanya bantuan dan dorongan dari berbagai pihak. Oleh karena itu pada kesempatan ini tidak lupa penulis mengucapkan terimakasih atas bantuan yang diberikan selama skripsi ini kepada yang terhormat :

1. Dr. Marzuki Alie, SE.,MM, selaku Rektor Universitas Indo Global Mandiri Palembang.
2. Rudi Heriansyah, ST., M.Eng. Ph.D selaku Dekan Fakultas Ilmu Komputer Universitas Indo Global Mandiri dan Dosen Pembimbing I.
3. Dewi Sartika, M.Kom, sebagai Ketua Program Studi Teknik Informatika Universitas Indo Global Mandiri.
4. Latri Widya Astuti, M.Kom, sebagai Dosen Pembimbing II.
5. Ir. Nazori Suhandi, MM, sebagai Pembimbing Akademik.
6. Bapak/Ibu Dosen Fakultas Ilmu Komputer dan Karyawan/Karyawati Universitas Indo Global Mandiri.
7. Kedua Orang Tua saya dan Keluarga yang telah memberikan dukungan moril maupun materil, serta Do'a sehingga penulis bisa menyelesaikan skripsi ini.
8. Semua teman-teman seperjuangan Teknik Informatika Angkatan 2019.



Dengan segala kerendahan hati penulis menyadari bahwa skripsi ini masih jauh dari kata sempurna, oleh karena itu dibutuhkan kritik dan saran untuk perbaikan dan pengembangan skripsi ini sangat diharapkan. Akhir kata, semoga skripsi ini bermanfaat bagi semua pihak terima kasih.

Palembang, Juli 2023

Penulis

Devi Yusrina

2019.11.0040

## DAFTAR ISI

<b>HALAMAN JUDUL LUAR .....</b>	<b>i</b>
<b>HALAMAN JUDUL DALAM .....</b>	<b>ii</b>
<b>LEMBAR PENGESAHAN SKRIPSI .....</b>	<b>iii</b>
<b>LEMBAR PERSETUJUAN DEWAN PENGUJI .....</b>	<b>iv</b>
<b>SURAT KETERANGAN REVISI.....</b>	<b>v</b>
<b>ABSTRAK .....</b>	<b>vi</b>
<b>ABSTRACT .....</b>	<b>vii</b>
<b>KATA PENGANTAR.....</b>	<b>viii</b>
<b>DAFTAR ISI.....</b>	<b>x</b>
<b>DAFTAR TABEL .....</b>	<b>xii</b>
<b>DAFTAR GAMBAR.....</b>	<b>xiii</b>
<b>DAFTAR LAMPIRAN .....</b>	<b>xiv</b>
<b>BAB 1 PENDAHULUAN .....</b>	<b>1</b>
1.1 Latar Belakang .....	1
1.2 Perumusan Masalah .....	2
1.3 Batasan Masalah.....	2
1.4 Tujuan Penelitian .....	2
1.5 Manfaat Penelitian .....	3
1.6 Sistematika Penulisan .....	3
<b>BAB 2 LANDASAN TEORI .....</b>	<b>5</b>
2.1 Plagiarisme.....	5
2.2 <i>Text Mining</i> .....	6
2.2.1 <i>Case Folding</i> .....	7
2.2.2 <i>Tokenizing</i> .....	7
2.2.3 <i>Stemming</i> .....	8
2.3 <i>Sentence Similarity Measures</i> .....	8
2.3.1 <i>Pengertian</i> .....	8
2.3.2 <i>Jaccard Similarity</i> .....	10
2.3.3 <i>Cosine Similarity</i> .....	11
2.4 <i>Confusion Matrix</i> .....	12

2.5	<i>Flowchart</i> .....	14
2.6	<i>Javascript</i> .....	16
2.7	Basis Data ( <i>Database</i> ) .....	17
2.8	<i>XAMPP</i> .....	18
<b>2.9</b>	<b>Penelitian Terdahulu</b> .....	<b>20</b>
<b>BAB 3 METODOLOGI PENELITIAN</b> .....		<b>29</b>
3.1	Tahapan Penelitian .....	29
3.2	Studi Literatur .....	30
3.3	Pengumpulan Data .....	30
3.4	Analisis Masalah .....	30
3.5	Perancangan Sistem .....	31
3.5.1	Kebutuhan <i>hardware</i> dan <i>software</i> .....	31
3.5.2	<i>Flowchart Jaccard Similarity</i> .....	32
3.5.3	<i>Flowchart Cosine Similarity</i> .....	35
3.5.4	Penerapan Algoritma <i>Jaccard Similarity</i> dan <i>Cosine Similarity</i> ....	38
3.6	Evaluasi .....	45
3.7	Kesimpulan .....	47
<b>BAB 4 HASIL DAN PEMBAHASAN</b> .....		<b>48</b>
4.1	Implementasi .....	48
4.1.1	Batasan Implementasi .....	48
4.1.2.	Implementasi Tampilan Antarmuka .....	48
4.2	Pengujian Program .....	50
4.2.1	Tujuan Pengujian .....	50
4.2.2	Data Pengujian .....	50
4.2.4	Hasil Pengujian .....	55
4.2.5	Visualisasi Hasil Pengujian .....	62
4.3	Evaluasi .....	66
4.3.1	Hasil Evaluasi .....	68
<b>BAB 5 KESIMPULAN DAN SARAN</b> .....		<b>70</b>
5.1	Kesimpulan .....	70
5.2	Saran .....	70
<b>DAFTAR PUSTAKA</b> .....		<b>71</b>

## DAFTAR TABEL

Tabel 2.1 Contoh <i>Case folding</i> .....	7
Tabel 2. 2 Contoh <i>Tokenizing</i> .....	7
Tabel 2.3 Contoh <i>Stemming</i> .....	8
Tabel 2.4 Contoh <i>Jaccard Similarity</i> .....	10
Tabel 2. 5 <i>Confusion Matrix Table</i> (Ainur Rohman & Arifianto, 2021).....	13
Tabel 2.6 Simbol <i>Flowchart</i> (Rosaly et al., n.d.).....	14
Tabel 2.7 Penelitian Terdahulu .....	20
Tabel 3.1 Daftar kebutuhan perangkat keras ( <i>hardware</i> ) .....	31
Tabel 3.2 Daftar kebutuhan perangkat lunak ( <i>software</i> ).....	31
Tabel 3.3 Penerapan algoritma <i>Jaccard Similarity 1</i> .....	38
Tabel 3.4 Penerapan algoritma <i>Jaccard Similarity 2</i> .....	40
Tabel 3.5 Penerapan algoritma <i>Jaccard Similarity 3</i> .....	40
Tabel 3.6 Penerapan algoritma <i>Cosine Similarity 1</i> .....	41
Tabel 3.7 Penerapan algoritma <i>Cosine Similarity 2</i> .....	43
Tabel 3.8 Penerapan algoritma <i>Cosine Similarity 3</i> .....	44
Tabel 3.9 Hasil Klasifikasi <i>Confusion Matrix Jaccard Similarity</i> .....	46
Tabel 3.10 Hasil Klasifikasi <i>Confusion Matrix Cosine Similarity</i> .....	46
Tabel 4.1 <i>Data Training</i> .....	51
Tabel 4.2 <i>Data Training</i> (lanjutan) .....	52
Tabel 4.3 <i>Data Training</i> (lanjutan) .....	53
Tabel 4.4 <i>Data Training</i> (lanjutan) .....	54
Tabel 4.5 <i>Data Testing</i> .....	54
Tabel 4.6 Rincian Skenario Pengujian .....	55
Tabel 4.7 Hasil Pengujian Setiap Skenario .....	56
Tabel 4.8 Data Hasil 5 Skenario Pengujian .....	66
Tabel 4.9 <i>Confusion matrix Cosine Similarity</i> .....	67
Tabel 4.10 <i>Confusion matrix Jaccard Similarity</i> .....	67

## DAFTAR GAMBAR

Gambar 2.1 <i>Text mining proses</i> (Fayyaz & Hadizadeh, 2020).....	6
Gambar 2.2 Pendekatan Semantik <i>Sentence Similarity Measures</i> (Farouk, 2019). 9	
Gambar 2.3 Visualisasi <i>Jaccard Similarity</i> .....	11
Gambar 3.1 Tahapan penelitian .....	29
Gambar 3.2 <i>Flowchart Jaccard Similarity</i> .....	32
Gambar 3.3 <i>Flowchart Text Preprocessing Jaccard</i> .....	33
Gambar 3.4 <i>Flowchart Jaccard Similarity measure</i> .....	34
Gambar 3.5 <i>Flowchart Cosine Similarity</i> .....	35
Gambar 3.6 <i>Flowchart Text Preprocessing Cosine</i> .....	36
Gambar 3.7 <i>Flowchart Cosine Similarity measure</i> .....	37
Gambar 4.1 Tampilan antarmuka deteksi plagiat .....	49
Gambar 4.2 Tampilan input file abstrak .....	50

## **DAFTAR LAMPIRAN**

Lampiran 1 : Daftar Riwayat Hidup

Lampiran 2 : Kartu Bimbingan Skripsi

Lampiran 3 : Surat Pernyataan Tidak Plagiat

Lampiran 4 : Pengujian 4 dan pengujian 5

Lampiran 5 : *Source Code*