

**PENERAPAN METODE DICE SIMILARITY COEFFICIENT UNTUK  
MENDETEKSI KEMIRIPAN ARTIKEL BERBASIS DESKTOP**

**SKRIPSI**

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana  
Program Studi Informatika



disusun oleh

**LALU MUHAMAD DWANDA FATANATIL ULUM**

**15.11.8919**

Kepada

**FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA**

**2022**

**PENERAPAN METODE DICE SIMILARITY COEFFICIENT  
UNTUK MENDETEKSI KEMIRIPAN ARTIKEL  
BERBASIS DESKTOP**

**SKRIPSI**

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana  
Program Studi Informatika



disusun oleh

**LALU MUHAMAD DWANDA FATANATIL ULUM  
15.11.8919**

Kepada

**FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA**

**2022**

**HALAMAN PERSETUJUAN**

**SKRIPSI**

**PENERAPAN METODE DICE SIMILARITY COEFFICIENT UNTUK  
MENDETEKSI KEMIRIPAN ARTIKEL BERBASIS DESKTOP**

yang disusun dan diajukan oleh

**Lalu Muhamad Dwanda Fatanatil Ulum**

**15.11.8919**

telah disetujui oleh Dosen Pembimbing Skripsi  
pada tanggal 26 Juli 2022

**Dosen Pembimbing,**

**Wiwi Widayani, M.Kom**

**NIK. 190302272**

**HALAMAN PENGESAHAN**

**SKRIPSI**

**PENERAPAN METODE DICE SIMILARITY COEFFICIENT UNTUK  
MENDETEKSI KEMIRIPAN ARTIKEL BERBASIS DESKTOP**

yang disusun dan diajukan oleh

**Lalu Muhamad Dwanda Fatanatil Ulum**

**15.11.8919**

Telah dipertahankan di depan Dewan Penguji  
pada tanggal 26 Juli 2022

**Susunan Dewan Penguji**

**Nama Penguji**

**Tanda Tangan**

**Mardhiya Hayaty, S.T., M.Kom**  
**NIK. 190302108**

**Supriatin, M.Kom**  
**NIK. 190302239**

**Wiwi Widayani, M.Kom**  
**NIK. 190302272**

Skripsi ini telah diterima sebagai salah satu persyaratan  
untuk memperoleh gelar Sarjana Komputer  
Tanggal 26 Juli 2022

**DEKAN FAKULTAS ILMU KOMPUTER**

**Hanif Al Fatta, S.Kom., M.Kom.**  
**NIK. 190302096**

## HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : **Lalu Muhamad Dwanda Fatanatil Ulum**  
NIM : **15.11.8919**

Menyatakan bahwa Skripsi dengan judul berikut:

### **Penerapan Metode Dice Similarity Coefficient Untuk Mendeteksi Kemiripan Artikel Berbasis Desktop**

Dosen Pembimbing : **Wiwi Widayani, M.Kom**

1. Karya tulis ini adalah benar-benar **ASLI** dan **BELUM PERNAH** diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian **SAYA** sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab **SAYA**, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini **SAYA** buat dengan **sesungguhnya**, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka **SAYA** bersedia menerima **SANKSI AKADEMIK** dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 26 Juli 2022

Yang Menyatakan,

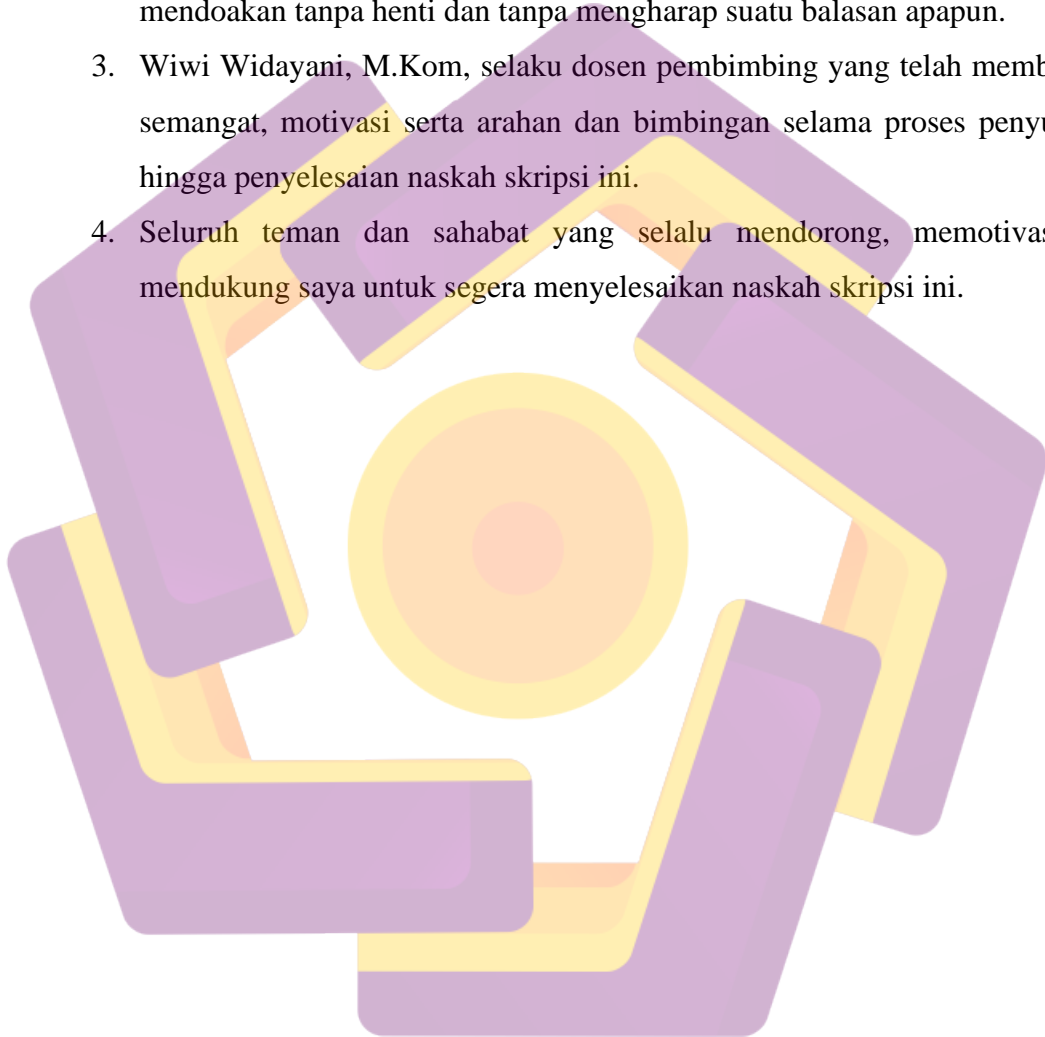


Lalu Muhamad Dwanda Fatanatil Ulum

## HALAMAN PERSEMBAHAN

Dengan mengucap puja puji syukur kehadiran Allah SWT, Alhamdulillah atas segala nikmat dan karunia-Nya. Skripsi ini saya persembahkan kepada:

1. Allah SWT, yang telah banyak memberikan karunia-Nya, salah satunya memberikan kemudahan dan kelancaran dalam proses penulisan naskah ini.
2. Kedua orang tua saya yang selalu membimbing, mendukung, memotivasi dan mendoakan tanpa henti dan tanpa mengharap suatu balasan apapun.
3. Wiwi Widayani, M.Kom, selaku dosen pembimbing yang telah memberikan semangat, motivasi serta arahan dan bimbingan selama proses penyusunan hingga penyelesaian naskah skripsi ini.
4. Seluruh teman dan sahabat yang selalu mendorong, memotivasi dan mendukung saya untuk segera menyelesaikan naskah skripsi ini.



## KATA PENGANTAR

Puji syukur penulis panjatkan kehadirat Allah, Tuhan yang maha Esa, atas segala rahmat, nikmat, kelancaran, kemudahan dan karunia ilmu sehingga penulis dapat menyelesaikan skripsi ini yang berjudul “Penerapan Metode Dice Similarity Coefficient Untuk Mendeteksi Kemiripan Artikel Berbasis Desktop”.

Skripsi ini disusun untuk memenuhi salah satu persyaratan kelulusan kelulusan di Program Strata-I Informatika di Universitas Amikom Yogyakarta. Banyak pihak yang telah mendukung terselesaikannya naskah skripsi ini, sehingga pada kesempatan ini penulis mengucapkan banyak terima kasih kepada:

1. Bapak Prof. Dr. M. Suyanto, MM. Selaku Rektor Universitas Amikom Yogyakarta.
2. Wiwi Widayani, M.Kom selaku dosen pembimbing yang selalu memberikan semangat, motivasi selama bimbingan dalam menyelesaikan skripsi ini.
3. Bapak dan Ibu Dosen Universitas Amikom Yogyakarta yang telah memberikan banyak ilmu yang sangat bermanfaat bagi saya.
4. Teman dan sahabat saya yang selalu menyemangati, memotivasi dan mengingatkan saya untuk menyelesaikan skripsi.

Penulis juga memohon maaf apabila dalam penyusunan skripsi ini masih banyak kekurangan dan masih jauh dari kata sempurna. Oleh karena itu penulis berhadap kepada semua pihak agar dapat menyampaikan kritik dan saran yang membangun untuk menambah kesempurnaan skripsi ini.

Yogyakarta, 26 Juli 2022

Penulis

## DAFTAR ISI

HALAMAN JUDUL .....	i
HALAMAN PERSETUJUAN .....	ii
HALAMAN PENGESAHAN .....	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI.....	iv
HALAMAN PERSEMBAHAN .....	v
KATA PENGANTAR .....	vi
DAFTAR ISI.....	vii
DAFTAR TABEL.....	x
DAFTAR GAMBAR.....	xi
INTISARI .....	xii
<i>ABSTRACT</i> .....	xiii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang Masalah.....	1
1.2 Rumusan Masalah .....	2
1.3 Batasan Masalah.....	3
1.4 Tujuan Penelitian .....	3
1.5 Manfaat Penelitian .....	3
1.6 Metode Penelitian.....	4
1.6.1 Metode Pengumpulan Data .....	4
1.6.2 Metode Analisis.....	4
1.6.3 Metode Perancangan .....	4
1.6.4 Metode Pengembangan .....	4
1.6.5 Metode Testing.....	5
1.7 Sistematika Penulisan.....	5
1.7.1 BAB I : Pendahuluan.....	5



1.7.2	BAB II : Landasan Teori .....	5
1.7.3	BAB III : Analisa dan Perancangan Sistem .....	5
1.7.4	BAB IV : Implementasi dan Pembahasan .....	5
1.7.5	BAB V : Penutup.....	5
<b>BAB II LANDASAN TEORI.....</b>		<b>6</b>
2.1	Tinjauan Pustaka .....	6
2.2	Information Retrieval .....	14
2.3	Text Preprocessing .....	14
2.3.1	Case Folding.....	14
2.3.2	Tokenizing .....	14
2.3.3	Stopword Filtering.....	15
2.4	Stemming .....	15
2.4.1	Algoritma Nazief dan Adriani .....	15
2.5	<i>Similarity</i> (Kemiripan) .....	20
2.6	Dice Similarity Coefficient .....	20
2.7	Flowchart.....	21
2.8	Black-box .....	23
2.9	C#.....	23
2.10	IDE (Integrated Development Environment) .....	23
2.10.1	Microsoft Visual Studio .....	24
2.11	WPF.....	24
2.12	Database .....	24
2.13	SQL.....	26
2.14	DBMS (Database Management System).....	26
2.15	SQLite.....	27
<b>BAB III ANALISIS DAN PERANCANGAN .....</b>		<b>28</b>
3.1	Tinjauan Umum .....	28

3.2	Analisis Kebutuhan .....	28
3.2.1	Kebutuhan Fungsional.....	28
3.2.2	Kebutuhan Non-Fungsional .....	29
3.3	Analisis Kelayakan Sistem.....	30
3.3.1	Kelayakan Teknis .....	30
3.3.2	Kelayakan Hukum .....	30
3.3.3	Kelayakan Operasional.....	30
3.4	Perancangan Sistem .....	31
3.4.1	Perancangan Proses .....	31
3.4.2	Perancangan <i>Database</i> .....	39
3.4.3	Perancangan <i>Interface</i> .....	40
<b>BAB IV IMPLEMENTASI DAN PEMBAHASAN .....</b>		<b>43</b>
4.1	Impelementasi (Produksi) .....	43
4.1.1	Membangun <i>Database</i> .....	43
4.1.2	Pembuatan User Interface.....	51
4.1.3	Koneksi <i>Database</i> .....	52
4.1.4	Implementasi Sistem .....	56
4.2	Pasca Produksi (Pengujian).....	66
4.2.1	Skenario Pengujian.....	66
4.2.2	Butir Pengujian.....	67
4.2.3	Hasil Uji .....	85
<b>BAB V PENUTUP .....</b>		<b>86</b>
5.1	Kesimpulan .....	86
5.2	Saran.....	86
<b>DAFTAR PUSTAKA .....</b>		<b>87</b>

## DAFTAR TABEL

Tabel 1.1 Perbandingan Hasil Stemming Porter dan Nazief & Adriani .....	2
Tabel 2.1 Perbandingan Penelitian .....	9
Tabel 2.2 Imbuan Yang Tidak Diizinkan .....	16
Tabel 2.3 Aturan Penghapusan Awalan Me- .....	17
Tabel 2.4 Aturan Penghapusan Awalan Pe- .....	18
Tabel 2.5 Aturan Penghapusan Awalan Be- .....	19
Tabel 2.6 Aturan Penghapusan Awalan Te- .....	19
Tabel 2.7 Simbol-simbol Flowchart .....	21
Tabel 3.1 Spesifikasi Perangkat Keras .....	29
Tabel 3.2 Contoh Case Folding .....	34
Tabel 3.3 Contoh Tokenizing .....	35
Tabel 3.4 Contoh Filtering Stopword .....	36
Tabel 4.1 Skenario Pengujian .....	66
Tabel 4.2 Pengujian Halaman Input Artikel .....	68
Tabel 4.3 Hasil Pengujian Halaman Hasil .....	68
Tabel 4.4 Pengujian Method Text Preprocessing Dan Stemming Nazief & Adriani .	69
Tabel 4.5 Pengujian Kemiripan Dengan Metode Filtering .....	71
Tabel 4.6 Pengujian Kemiripan Tanpa Metode Filtering .....	77

## DAFTAR GAMBAR

Gambar 3.1 Use Case User .....	31
Gambar 3.2 Flowchart Sistem .....	32
Gambar 3.3 Flowchart Preprocessing Text.....	33
Gambar 3.4 Flowchart Case Folding .....	34
Gambar 3.5 Flowchart Filtering.....	36
Gambar 3.6 Flowchart Stemming Algoritma Nazief dan Adriani.....	37
Gambar 3.7 Flowchart Dice Similarity Coefficient.....	39
Gambar 3.8 Struktur Tabel katadasar .....	39
Gambar 3.9 Struktur Tabel Stopword.....	40
Gambar 3.10 Interface Input Artikel.....	41
Gambar 3.11 Interface Hasil Similarity.....	42
Gambar 4.1 Membuat Database.....	44
Gambar 4.2 Tambah Database.....	44
Gambar 4.3 Simpan Database.....	45
Gambar 4.4 Database Berhasil Dibuat.....	46
Gambar 4.5 Open SQL & editor.....	47
Gambar 4.6 Query Pembuatan Tabel.....	48
Gambar 4.7 Query Execution Berhasil.....	49
Gambar 4.8 Import Ke Database .....	50
Gambar 4.9 Pilih File .CSV .....	50
Gambar 4.10 Import Data Berhasil.....	51
Gambar 4.11 Tampilan Input Artikel .....	51
Gambar 4.12 Tampilan Hasil.....	52
Gambar 4.13 Buka Nuget Package .....	53
Gambar 4.14 Instalasi System.Data.SQLite.EF6.....	54

## INTISARI

Dewasa ini, informasi yang kita lihat di internet khususnya artikel ilmiah sering kali mempunyai struktur kata atau penulisan yang mirip dengan artikel ilmiah yang lain. Sebagai contoh adalah artikel-artikel ilmiah yang sering kita jumpai di internet. Hal ini menyebabkan rancunya sumber informasi sehingga pembaca bingung akan sumber informasi asal. Begitupun dengan para penulis artikel ilmiah yang tidak ingin karya tulisnya mirip dengan artikel ilmiah yang lain sehingga penulis artikel ilmiah tidak melakukan tindakan plagiarisme dengan tidak disengaja.

Berdasarkan permasalahan di atas peneliti ingin membuat aplikasi untuk menguji kemiripan artikel, maka dibutuhkan algoritma pemroses kata seperti Algoritma Nazief & Adriani. Dalam Algoritma Nazief & Adriani terdapat sebuah proses stemming yaitu proses mengubah kata-kata di dalam sebuah dokumen teks ke bentuk kata dasarnya. Stemming Algoritma Nazief & Adriani dapat menambah keakuratan pencocokan kata, dalam prosesnya terdapat proses filtering yang khusus digunakan untuk mengurangi kata-kata tidak penting yang akan di-stemming.

Aplikasi yang dihasilkan berupa file desktop ".exe" yang ditunjukkan untuk memberikan gambaran atau hasil penelitian penerapan proses stemming Algoritma Nazief & Adriani.

**Kata Kunci:** Algoritma, Nazief & Adriani, Stemming, Filtering, Artikel Ilmiah.

## **ABSTRACT**

*Today, the information we see on the internet, especially scientific articles, often has a word structure or writing that is similar to other scientific articles. An example is the scientific articles that we often encounter on the internet. This causes confusion in the source of information so that readers are confused about the source of the original information. Likewise with the writers of scientific articles who do not want their writings to be similar to other scientific articles so that the authors of scientific articles do not commit acts of plagiarism unintentionally.*

*Based on the problems above, the researcher wants to make an application to test the similarity of articles, so a word processing algorithm is needed such as the Nazief & Adriani Alogritma. In the Nazief & Adriani Algorithm there is a stemming process, namely the process of changing the words in a text document to their basic word form. Stemming The Nazief & Adriani algorithm can increase the accuracy of word matching, in the process there is a special filtering process used to reduce unimportant words that will be stemmed.*

*The resulting application is a desktop file ".exe" which is shown to provide an overview or research results on the application of the Nazief & Adriani Algorithm stemming process.*

**Keywords:** *Algorithm, Nazief & Adriani, Stemming, Filtering, Scientific Articles.*