

**PENERAPAN METODE TEXT MINING PADA APLIKASI PENDETEKSI**

**KESAMAAN JUDUL TUGAS AKHIR PADA MAHASISWA**

(Studi Kasus : Program Studi Ilmu Komputer

Universitas Katolik Widya Mandira Kupang)

**TUGAS AKHIR**

**NO.982/WM.FT.H6/T.ILKOM/TA/2023**

**Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh**

**Gelar Sarjana Komputer**



Disusun Oleh:

**FRENGKY NUNES CABRAL**

**23119055**

**PROGRAM STUDI ILMU KOMPUTER**

**FAKULTAS TEKNIK**

**UNIVERSITAS KATOLIK WIDYA MANDIRA**

**2023**

HALAMAN PERSETUJUAN  
TUGAS AKHIR  
NO.982/WM.FT.H6/T.ILKOM/TA/2023

PENERAPAN METODE TEXT MINING PADA APLIKASI PENDETEKSI  
KESAMAAN JUDUL TUGAS AKHIR PADA MAHASISWA  
(Studi Kasus : Program Studi Ilmu Komputer  
Universitas Katolik Widya Mandira Kupang)


Oleh :

**FRENGKY NUNES CABRAL**

NIM : 23119055

TELAH DIPERIKSA/DISETUJUI OLEH PENGUJI  
DI : KOTA KUPANG  
PADA TANGGAL : AGUSTUS 2023

**DOSEN PENGUJI I**

  
**Paskalis A. Nani, S.T., M.T**  
NIDN : 0831038602

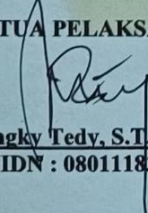
**DOSEN PENGUJI II**

  
**Yovinia C. Hoar Siki, S.T., M.T**  
NIDN : 0805058803


**DOSEN PENGUJI III**

  
**Frengky Tedy, S.T., M.T**  
NIDN : 0801118302

**KETUA PELAKSANA**

  
**Frengky Tedy, S.T., M.T**  
NIDN : 0801118302

**SEKRETARIS PELAKSANA**

  
**Ign. Pricher A.N. Samane, S.Si., M.Eng**  
NIDN : 0818098102

HALAMAN PENGESAHAN

TUGAS AKHIR

NO.982/WM.FT.H6/T.ILKOM/TA/2023

PENERAPAN METODE TEXT MINING PADA APLIKASI PENDETEKSI

KESAMAAN JUDUL TUGAS AKHIR PADA MAHASISWA

(Studi Kasus : Program Studi Ilmu Komputer

Universitas Katolik Widya Mandira Kupang)

Oleh:

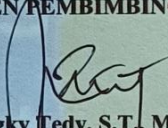
**FRENGKY NUNES CABRAL**

23119055

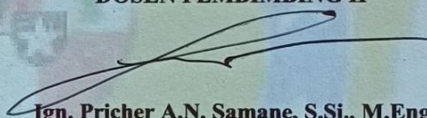
TELAH DIPERTAHANKAN DIDEPAN PEMBIMBING:

DOSEN PEMBIMBING I

DOSEN PEMBIMBING II

  
**Frengky Tedy, S.T., M.T**

NIDN : 0801118302

  
**Ign. Pricher A.N. Samane, S.Si., M.Eng**

NIDN : 0818098102

MENGETAHUI,

MENGESAHKAN,

KETUA PROGRAM STUDI

DEKAN FAKULTAS TEKNIK


ILMU KOMPUTER

UNIKA WIDYA MANDIRA

UNIKA WIDYA MANDIRA

  
**Sisilia D. Bakka Mau, S.Kom., M.T**

NIDN : 0807098502

  
**Dr. Don Gaspar N. Da Costa, S.T., M.T**

NIDN : 0820036801

## HALAMAN PERSEMBAHAN

Karya ini secara khusus saya persembahkan untuk:

TUHAN YESUS DAN BUNDA MARIA.

Untuk mereka yang memupuk mimpi dan mendukung perjuanganku,

Untuk keluargaku tercinta dan khususnya untuk Almarhumah Mama dan teman-temanku dari ujung Timur, Timor-leste hingga Kupang NTT yang selalu ada di sampingku,

Kata persembahan ini kuhaturkan dengan penuh terima kasih,

Karena tanpa kalian, skripsi ini takkan pernah terwujud.

Semoga hasil karya ini dapat memberikan manfaat yang sebesar-besarnya.

Terima kasih atas segala dukungan dan cinta kasih yang telah kalian berikan.

Teman-teman Angkatan 2019 terkasih

dan

Almamater UNWIRA Tercinta.

MOTTO

**DOBRAK KEMUNKINAN YANG  
TERBURUK BAGAIKAN ANAK PANAH  
MENEROBOS ANGIN  
HINGGA TITIK YANG  
DIINGINKAN.**

## PERNYATAAN KEASLIAN KARYA

Yang bertandatangan dibawah ini :

Nama : Frengky Nunes Cabral

No.Registrasi : 23119055

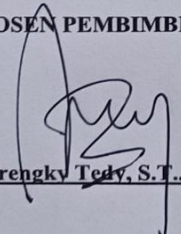
Fakultas : Teknik

Program Studi : Ilmu Komputer

Menyatakan bahwa, karya tulis skripsi dengan judul **“PENERAPAN METODE TEXT MINING PADA APLIKASI PENDETEKSI KESAMAAN JUDUL TUGAS AKHIR PADA MAHASISWA (Studi Kasus : Program Studi Ilmu Komputer Universitas Katolik Widya Mandira Kupang)”** adalah benar-benar karya saya sendiri. Apabila dikemudian hari ditemukan bahwa saya melakukan tindakan plagiat maka saya akan menerima sanksi yang telah ditetapkan.

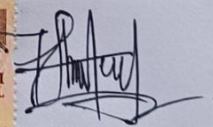
Disahkan/Diketahui

**DOSEN PEMBIMBING I**

  
**Frengky Tedv, S.T., M.T**

Kupang, Agustus 2023

**MAHASISWA/PEMILIK**

  
**Frengky Nunes Cabral**



## ABSTRAK

Program Studi Ilmu Komputer merupakan salah satu program studi yang berada di bawah naungan Fakultas Teknik, Universitas Katolik Widya Mandira Kupang. Penelitian ini bertujuan untuk merancang dan mengimplementasikan sebuah aplikasi *website* dengan menggunakan metode *text mining* dan algoritma *rabin-karp* untuk mendeteksi tingkat kesamaan judul dan abstrak tugas akhir yang diajukan oleh mahasiswa dan mahasiswi Program Studi Ilmu Komputer. Saat ini, proses pengajuan judul tugas akhir masih dilakukan secara manual, menyebabkan potensi adanya kesamaan judul dan abstrak antara mahasiswa dan mahasiswi. Untuk menjawab masalah tersebut metode *text mining* akan digunakan untuk menemukan pola yang menarik dari data tekstual yang berjumlah besar pada judul dan abstrak tugas akhir mahasiswa, serta algoritma *rabin-karp* yang digunakan untuk mencocokkan *string* judul dan abstrak tugas akhir yang diajukan dengan judul dan abstrak tugas akhir terdahulu agar dapat mengidentifikasi persentase tingkat kemiripan antara judul dan abstrak tugas akhir mahasiswa. Hasil dari penelitian ini adalah sebuah aplikasi yang dapat mendeteksi tingkat kesamaan atau similarity judul dan abstrak tugas akhir yang diajukan oleh setiap mahasiswa dan mahasiswi pada Program Studi Ilmu Komputer dengan ambang batas kemiripan yang telah ditentukan.

Kata kunci : Program Studi Ilmu Komputer, *Website Text Mining*, *Rabin-Karp*, *Similarity*.

## **ABSTRACT**

*The Computer Science Study Program is one of the study programs under the Faculty of Engineering, Widya Mandira Catholic University Kupang. This research aims to design and implement a website application using the text mining method and the Rabin-Karp algorithm to detect the similarity level of titles and abstracts of final projects submitted by male and female students of the Computer Science Study Program. Currently, the process of submitting final project titles is still done manually, leading to the potential for similarity in titles and abstracts among male and female students. To address this issue, the text mining method will be used to discover interesting patterns from the large textual data in the titles and abstracts of students' final projects, and the Rabin-Karp algorithm will be used to match the submitted title and abstract strings with previous final project titles and abstracts to identify the percentage of similarity between the titles and abstracts of students' final projects. The result of this research is an application that can detect the level of similarity or similarity of titles and abstracts of final projects submitted by each male and female student in the Computer Science Study Program with a predetermined threshold of similarity.*

*Keywords: Computer Science Study Program, Text Mining Website, Rabin-Karp, Similarity.*



## KATA PENGANTAR

Puji syukur ke hadirat Tuhan Yang Maha Esa atas berkat dan tuntunan-Nya yang tak henti sehingga penulisan Tugas Akhir ini dapat berjalan dengan baik. Adapun penulisan ini dibuat sebagai salah satu syarat untuk memenuhi dan memperoleh nilai Tugas Akhir.

Penulisan Tugas Akhir ini dapat berjalan dengan baik berkat adanya dukungan dari banyak pihak, baik berupa dukungan moral maupun materi. Untuk itu pada kesempatan ini penulis hendak menyampaikan ucapan terima kasih yang tak terhingga kepada:

1. Pater Dr. Philipus Tule, SVD selaku Rektor Universitas Katolik Widya Mandira Kupang.
2. Bapak Dr. Don Gaspar N. da Costa, S.T., M.T selaku Dekan Fakultas Teknik Universitas Katolik Widya Mandira Kupang.
3. Ibu Sisilia Daeng Bakka Mau, S.Kom., MT selaku Ketua Program Studi Ilmu Komputer
4. Bapak Frengky Tedy, S.T., M.T & Bapak Ignatius Pricher A.N. Samane, S.Si., M.Eng selaku dosen pembimbing I dan II. Terima kasih untuk waktu, pemikiran dan kesabaran yang diberikan selama proses bimbingan tugas akhir ini.
5. Bapak Paskalis Andrianus Nani, S.T., M.T & Ibu Yovinia Carmeneja Hoar Siki, S.T., M.T selaku dosen penguji I dan II, yang telah meluangkan waktu, tenaga dan pikiran dalam mengarahkan penulis untuk menyelesaikan tugas akhir.
6. Bapak Frengky Tedy, S.T., M.T selaku dosen pembimbing akademik yang selalu memberikan motivasi dan dorongan.
7. Pada semua Dosen dan karyawan Program Studi Teknik Ilmu Komputer.
8. Kedua orang tua tercinta, Bapak Justino Nunes dan Almarhumah Mama

Lucinda Pereira Marquês dan semua Kakak-kakakku serta Keponakan dan semua keluarga yang telah memberikan semangat dan dukungan.

9. Kakak Adino dan Kakak Jenny, Kakak Elvis dan Kakak Akita, Kakak Arif dan Kakak Anita, yang selalu mendukung, mengmotvasi dan memberikan kecerahan serta semangat.
10. Ditha B Mau yang selalu memberikan doa, dukungan dan semangat.
11. Sahabat – sahabatku Glanter, Oman, Jecky, Maya, Andre TMK dan yang telah berjuang bersama dan yang telah membantu dari awal perkuliahan hingga selesai pada Program Studi Teknik Ilmu Komputer, khususnya untuk angkatan 2019 tercinta.
12. Seluruh pihak yang telah memberikan sumbangsih dalam penyelesaian Tugas Akhir ini yang tidak dapat disebutkan namanya satu persatu, Tuhan Yesus kiranya membalas budi baik saudara – saudari sekalian.

Penulis menyadari dalam penyusunan Tugas Akhir ini, masih terdapat banyak kekurangan dan kelemahan yang dimiliki penulis baik itu sistematika penulisan maupun penggunaan bahasa. Oleh karena itu, penulis mengharapkan kritik dan saran dari berbagai pihak yang bersifat membangun. Semoga Tugas Akhir ini berguna bagi pembaca secara umum dan penulis secara khusus untuk kemajuan serta perkembangan. Akhir kata, penulis mengucapkan banyak terima kasih.

Kupang, Agustus 2023

Penulis

## DAFTAR ISI

<b>HALAMAN JUDUL .....</b>	<b>i</b>
<b>HALAMAN PERSETUJUAN .....</b>	<b>ii</b>
<b>HALAMAN PENGESAHAN .....</b>	<b>iii</b>
<b>HALAMAN PERSEMBAHAN .....</b>	<b>iv</b>
<b>MOTTO .....</b>	<b>v</b>
<b>PERNYATAAN KEASLIAN KARYA .....</b>	<b>vi</b>
<b>ABSTRAK .....</b>	<b>vii</b>
<b>ABSTRACT .....</b>	<b>viii</b>
<b>KATA PENGANTAR .....</b>	<b>ix</b>
<b>DAFTAR ISI .....</b>	<b>xi</b>
<b>DAFTAR TABEL .....</b>	<b>xv</b>
<b>DAFTAR GAMBAR .....</b>	<b>xvii</b>
<b>BAB I PENDAHULUAN .....</b>	<b>1</b>
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah .....	3
1.3 Batasan Masalah .....	3
1.4 Tujuan Penelitian .....	3

1.5	Manfaat Penelitian .....	4
1.6	Metodologi Penelitian .....	4
1.6.1	<i>Analysis</i> .....	5
1.6.2	<i>Design</i> .....	5
1.6.3	<i>Implementation</i> .....	6
1.6.4	<i>Testing / Verification</i> .....	6
1.6.5	<i>Maintenance</i> .....	7
1.7	Sistematika Penulisan .....	7
<b>BAB II LANDASAN TEORI .....</b>		<b>9</b>
2.1	Penelitian Sebelumnya .....	9
2.2	Gambaran Umum Program studi ilmu komputer .....	16
2.2.1	Sejarah Program Studi Ilmu Komputer .....	16
2.2.2	Tujuan dari aplikasi deteksi kesamaan judul mahasiswa .....	18
2.3	Teori Penunjang.....	18
2.3.1	<i>Data Mining</i> .....	18
2.3.2	<i>Text Mining</i> .....	19
2.3.3	<i>Text Preprocessing</i> .....	19
2.3.4	Algoritma <i>Rabin-Karp</i> .....	20

2.3.5	PHP ( <i>Hypertext Preprocessor</i> ) .....	24
2.3.6	<i>JavaScript</i> .....	25
2.3.7	<i>MySQL</i> .....	25
2.3.8	<i>Web</i> .....	25
2.3.9	<i>MVC (Model View Controller)</i> .....	26
2.4	Perancangan Sistem .....	27
2.4.1	Antarmuka Perancangan .....	27
<b>BAB III ANALISIS DAN PERANCANGAN SISTEM .....</b>		<b>33</b>
3.1	Analisis Sistem .....	34
3.1.1	Analisis Kebutuhan Sistem .....	34
3.1.2	Analisis Peran Sistem .....	36
3.1.3	Kategori Pengguna Sistem .....	37
3.2	Perancangan Sistem .....	37
3.2.1	<i>Flowchart</i> Sisten Berjalan .....	37
3.2.2	<i>Flowchart</i> Sistem .....	39
3.2.3	Diagram Konteks .....	40
3.2.4	Diagram Berjenjang (HIPO) .....	41
3.2.5	<i>Data Flow Diagram (DFD)</i> Level 1 .....	43

3.2.6	<i>Entity Relationship Diagram (ERD)</i> .....	44
3.2.7	Relasi Antar Tabel .....	45
3.3	Perancangan Tabel .....	45
3.4	Perancangan Antar Muka.....	50
<b>BAB IV IMPLEMENTASI SISTEM .....</b>		<b>58</b>
4.2	Implementasi Sistem .....	62
<b>BAB V PENGUJIAN DAN ANALISIS HASIL .....</b>		<b>78</b>
5.1	Pengujian Sistem .....	78
5.2	Analisis Hasil .....	80
5.3	Cara Kerja Manual .....	82
5.3.1	Langkah-langkah Awal .....	82
<b>BAB VI PENUTUP .....</b>		<b>97</b>
6.1	Kesimpulan .....	98
6.2	Saran .....	99
<b>DAFTAR PUSTAKA .....</b>		<b>99</b>

## DAFTAR TABEL

Tabel 2. 1 Perbandingan Terhadap Penelitian Sebelumnya .....	11
Tabel 2. 2 Contoh K-Gram .....	21
Tabel 2. 3 Contoh <i>Rolling Hash</i> .....	22
Tabel 2. 4 Simbol-Simbol Flowchart .....	28
Tabel 2. 5 Simbol-Simbol DFD .....	31
Tabel 2. 6 Simbol-Simbol ERD .....	32
Tabel 3. 1 <i>Admin</i> .....	46
Tabel 3. 2 abstrak_uji .....	46
Tabel 3. 3 Abstrak_asli .....	47
Tabel 3. 4 hasil_uji .....	47
Tabel 3. 5 Pengaturan .....	48
Tabel 3. 6 dokumen .....	48
Tabel 3. 7 abstrak_encode .....	49
Tabel 3. 8 v_uji .....	49
Tabel 5. 1 Tabel Pengujian .....	79
Tabel 5. 2 File Teks .....	82

Tabel 5. 3 <i>Text Preprocessing</i> .....	83
Tabel 5. 4 Hasil <i>Text Preprocessing</i> .....	85
Tabel 5. 5 Proses <i>K-Gram</i> .....	85
Tabel 5. 6 Hasil Proses <i>Hasing</i> .....	86
Tabel 5. 7 Hasil <i>Hasing</i> .....	90
Tabel 5. 8 Hasil <i>Fingerprint</i> .....	92
Tabel 5. 9 Jumlah Total <i>Fingerprint</i> .....	95



## DAFTAR GAMBAR

Gambar 1. 1 Tahap Penelitian . . . . .	5
Gambar 2. 2 <i>Relasi One To One</i> . . . . .	30
Gambar 2. 3 <i>Relasi One To Many</i> . . . . .	30
Gambar 2. 4 <i>Relasi Many To Many</i> . . . . .	31
Gambar 3. 1 <i>Flowchart</i> Sistem Berjalan . . . . .	39
Gambar 3. 2 <i>Flowchart</i> Sistem . . . . .	40
Gambar 3. 3 <i>Diagram Konteks</i> . . . . .	41
Gambar 3. 4 Diagram Berjenjang . . . . .	42
Gambar 3. 5 <i>Data Flow Diagram</i> (DFD) level 1 . . . . .	43
Gambar 3. 6 ERD . . . . .	44
Gambar 3. 7 Relasi Antar Tabel . . . . .	45
Gambar 3. 8 Desain Halaman Awal Pengguna . . . . .	51
Gambar 3. 9 Desain Halaman Pengecekan . . . . .	52
Gambar 3. 10 Hasil Cek Tingkat Plagiarisme . . . . .	53
Gambar 3. 12 <i>Dashboard Admin</i> . . . . .	54
Gambar 3. 13 Arsip Data . . . . .	55
Gambar 3. 14 Data Pengajuan Baru. . . . .	55
Gambar 3. 15 Hasil Uji Plagiarisme . . . . .	56
Gambar 3. 16 <i>Admin Profile</i> . . . . .	57
Gambar 3. 17 Pengaturan . . . . .	57

Gambar 4. 1 Tabel abstrak_asli. ....	58
Gambar 4. 2 Tabel abstrak_uji. ....	59
Gambar 4. 3 Tabel <i>admin</i> . ....	59
Gambar 4. 4 Tabel dokumen .....	60
Gambar 4. 5 Tabel hasil_uji. ....	60
Gambar 4. 6 Tabel pengaturan. ....	61
Gambar 4. 7 Tabel abstrak_encode. ....	61
Gambar 4. 8 Tabel v_uji.....	62
Gambar 4. 9 Halaman Awal Pengguna .....	63
Gambar 4. 10 Halaman Cek Judul .....	64
Gambar 4. 11 Halaman Hasil <i>Symilarity</i> .....	66
Gambar 4. 12 Halaman <i>Login Admin</i> .....	68
Gambar 4. 13 Halaman <i>Admin</i> .....	69
Gambar 4. 14 Halaman Arsip Data .....	71
Gambar 4. 15 Halaman Data Pengajuan Baru .....	73
Gambar 4. 16 Halaman hasil uji plagiarisme .....	74
Gambar 4. 17 Halaman <i>admin</i> .....	75
Gambar 4. 18 Halaman pengaturan .....	76