

**Penerapan Metode *TF-IDF* dan *Cosine Similarity* dalam Sistem
Web Pendeteksi Kemiripan Judul Tugas Akhir Terintegrasi
Repositori Digital di Prodi STMIK Palangkaraya**

PROPOSAL TUGAS AKHIR

Diajukan Untuk Memenuhi Salah Satu Syarat Penulisan Tugas Akhir
pada Sekolah Tinggi Manajemen Informatika Dan Komputer
(STMIK) Palangkaraya



OLEH

IBNU ROFIK

C2255201109

PROGRAM STUDI TEKNIK INFORMATIKA

**SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER
(STMIK) PALANGKARAYA
2025**

**Penerapan Metode *TF-IDF* dan *Cosine Similarity* dalam Sistem
Web Pendeteksi Kemiripan Judul Tugas Akhir Terintegrasi
Repositori Digital di Prodi STMIK Palangkaraya**

PROPOSAL TUGAS AKHIR

Diajukan Untuk Memenuhi Salah Satu Syarat Penulisan Tugas Akhir
pada Sekolah Tinggi Manajemen Informatika Dan Komputer
(STMIK) Palangkaraya

OLEH

IBNU ROFIK

C2255201109

PROGRAM STUDI TEKNIK INFORMATIKA

**SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER
(STMIK) PALANGKARAYA
2025**

PERSETUJUAN

Penerapan Metode *TF-IDF* dan *Cosine Similarity* dalam Sistem Web Pendeteksi Kemiripan Judul Tugas Akhir Terintegrasi Repositori Digital di Prodi STMIK Palangkaraya

Proposal Tugas Akhir Ini Telah Disetujui dan Disahkan

Pembimbing I,

Pembimbing II,

Dr. Maura Widyaningsih, M.Kom.
NIK. 197502232009002

Ir. Abdul Hadi, S.T., M.Kom.
NIK. 198505102021101

Mengetahui

Ketua STMIK Palangkaraya,

Ir. Hj. Siti Maryamah, MM.
NIK. 196902021995004

PENGESAHAN

Penerapan Metode *TF-IDF* dan *Cosine Similarity* dalam Sistem Web Pendeteksi Kemiripan Judul Tugas Akhir Terintegrasi Repository Digital di Prodi STMIK Palangkaraya

Proposal Tugas Akhir ini telah Diseminarkan, Dinilai, dan Disahkan
Oleh Tim Seminar pada Tanggal 7 Januari 2021

Tim Seminar Proposal :

1. Nama Ketua Penguji
Ketua
2. Nama Sekretaris Penguji
Sekretaris
3. Nama Anggota Penguji
Anggota

DAFTAR ISI

PERSETUJUAN	i
PENGESAHAN	ii
DAFTAR TABEL	iv
DAFTAR GAMBAR	v
DAFTAR LAMPIRAN	vi
BAB I	1
PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Perumusan Masalah	2
1.3 Batasan Masalah	3
1.4 Tujuan dan Manfaat	4
1.5 Sistematika Penulisan	5
BAB II	7
TINJAUAN PUSTAKA	7
2.1 Kajian Teori	7
2.2 Penelitian Yang Relevan	8
BAB III	10
TINJAUAN PUSTAKA	10
3.1 Tinjauan Umum	10
3.2 Jenis penelitian	10
3.3 Desain Penelitian	10
3.4 Instrumen Penelitian	10
3.5 dan seterusnya	10
DAFTAR PUSTAKA	11

DAFTAR TABEL

Tabel 2.1. Perbandingan penelitian6
------------------------------------	-------

DAFTAR GAMBAR

Gambar 2.1. Tahapan *Multimedia Development Life Cycle*

8

DAFTAR LAMPIRAN

- Lampiran 1. Surat tugas pembimbing tugas akhir
- Lampiran 2. Lembar konsultasi bimbingan Tugas Akhir
- Lampiran 3. Surat izin penelitian
- Lampiran 4. Surat keterangan telah melakukan penelitian
- Lampiran 5. Lembar wawancara
- Lampiran 6. Lembar dokumentasi wawancara
- Lampiran 7. Surat tugas penguji seminar
- Lampiran 8. Berita acara penilaian seminar proposal TA
dst

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Penyusunan tugas akhir merupakan salah satu syarat kelulusan bagi mahasiswa Program Studi Teknik Informatika STMIK Palangka Raya. Dalam proses pengajuan judul tugas akhir, sering terjadi permasalahan terkait kemiripan atau kesamaan judul antara mahasiswa. Kemiripan judul dapat menyebabkan duplikasi penelitian, mengurangi orisinalitas karya ilmiah, dan berpotensi menimbulkan tumpang tindih substansi penelitian.

Hingga saat ini, proses pengecekan kemiripan judul tugas akhir di Program Studi Teknik Informatika STMIK Palangka Raya masih dilakukan secara manual oleh dosen pembimbing atau koordinator tugas akhir. Belum tersedia sistem terintegrasi untuk pengecekan kesamaan judul sekaligus akses terhadap dokumen tugas akhir yang telah ada. Kondisi ini menyebabkan proses verifikasi menjadi kurang optimal, memerlukan waktu yang lama, serta menyulitkan mahasiswa dalam mengakses referensi penelitian terdahulu.

Seiring meningkatnya jumlah mahasiswa setiap tahunnya, volume data judul tugas akhir yang harus dicek semakin bertambah. Hal ini membuat proses pengecekan manual semakin tidak efisien dan rentan terhadap human error. Dibutuhkan solusi teknologi yang dapat mengotomatisasi proses pengecekan kemiripan judul secara cepat dan akurat, sekaligus menyediakan repositori digital untuk dokumen tugas akhir.

Untuk mengatasi permasalahan tersebut, diperlukan sistem berbasis web yang dapat mendeteksi kemiripan judul tugas akhir secara otomatis menggunakan algoritma Term Frequency-Inverse Document Frequency (TF-IDF) dan Cosine Similarity. TF-IDF merupakan metode pembobotan kata yang dapat mengidentifikasi tingkat kepentingan suatu kata dalam dokumen, sedangkan Cosine Similarity mengukur tingkat kesamaan antara dua dokumen berdasarkan sudut cosinus dari vektor yang terbentuk. Sistem ini juga akan menyediakan fitur repositori digital yang memungkinkan mahasiswa mengunggah dokumen proposal mereka setelah selesai seminar atau revisi, sehingga dapat menjadi referensi bagi mahasiswa lainnya.

Penerapan sistem ini diharapkan dapat meningkatkan efisiensi proses pengecekan judul, mengurangi duplikasi penelitian, dan memudahkan akses terhadap dokumen tugas akhir terdahulu di Program Studi Teknik Informatika STMIK Palangka Raya.

1.2 Perumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, maka rumusan masalah dalam penelitian ini adalah:

1. Bagaimana merancang dan membangun sistem berbasis web untuk mendeteksi kemiripan judul tugas akhir yang terintegrasi dengan repositori dokumen di Program Studi Teknik Informatika STMIK Palangka Raya?
2. Bagaimana mengimplementasikan algoritma TF-IDF dan Cosine Similarity dalam mendeteksi tingkat kemiripan judul tugas akhir secara akurat?

3. Bagaimana merancang fitur repositori digital yang memungkinkan penyimpanan, pengelolaan, dan akses dokumen tugas akhir mahasiswa?
4. Bagaimana sistem ini dapat meningkatkan efisiensi proses pengecekan judul dan mengurangi potensi duplikasi tugas akhir di Program Studi Teknik Informatika STMIK Palangka Raya?

1.3 Batasan Masalah

Agar penelitian ini lebih terarah dan fokus, maka ditetapkan batasan masalah sebagai berikut:

1. Sistem yang dibangun berbasis web dan hanya digunakan untuk mendeteksi kemiripan judul tugas akhir serta menyimpan metadata dan link dokumen tugas akhir di Program Studi Teknik Informatika STMIK Palangka Raya.
2. Data judul tugas akhir yang digunakan adalah data judul yang telah terdaftar dan tersimpan dalam database Program Studi Teknik Informatika STMIK Palangka Raya.
3. Algoritma yang digunakan untuk mendeteksi kemiripan judul adalah Term Frequency-Inverse Document Frequency (TF-IDF) dan Cosine Similarity.
4. Sistem mendeteksi kemiripan berdasarkan teks judul tugas akhir, tidak menganalisis isi lengkap dari dokumen tugas akhir.
5. Sistem menampilkan hasil berupa persentase kemiripan, daftar judul yang mirip, dan menyediakan akses untuk melihat dokumen proposal melalui embed link Google Drive.
6. Penyimpanan file PDF dilakukan di Google Drive milik mahasiswa masing-masing, sistem hanya menyimpan link URL untuk akses dokumen.

7. Input data tugas akhir (judul, penulis, tahun, jurusan, keywords, dan link PDF) dilakukan melalui Google Apps Script yang terintegrasi dengan Google Sheets.

1.4 Tujuan dan Manfaat

Tujuan dari penelitian ini adalah:

1. Merancang dan membangun sistem berbasis web untuk mendeteksi kemiripan judul tugas akhir yang terintegrasi dengan repositori dokumen di Program Studi Teknik Informatika STMIK Palangka Raya.
2. Mengimplementasikan algoritma TF-IDF dan Cosine Similarity dalam sistem pendeteksi kemiripan judul tugas akhir.
3. Mengembangkan fitur repositori digital yang memungkinkan mahasiswa mengunggah, menyimpan, dan mengakses dokumen tugas akhir dalam format PDF.
4. Menganalisis peningkatan efisiensi sistem dan pengurangan potensi duplikasi judul tugas akhir di Program Studi Teknik Informatika STMIK Palangka Raya.

Manfaat yang diharapkan dari penelitian ini adalah:

Bagi institusi:

1. Menyediakan sistem terintegrasi untuk manajemen data tugas akhir mahasiswa.
2. Meningkatkan kualitas dan orisinalitas karya ilmiah mahasiswa melalui pencegahan duplikasi penelitian.

3. Membangun repositori digital tugas akhir yang terorganisir dan mudah diakses.

Bagi Mahasiswa:

1. Memudahkan mahasiswa dalam mengecek kemiripan judul secara mandiri, cepat, dan real-time sebelum mengajukan proposal.
2. Menyediakan akses mudah untuk mempelajari penelitian terdahulu sebagai referensi dan inspirasi.
3. Membantu mahasiswa menghindari duplikasi penelitian secara tidak sengaja.

Bagi Dosen:

1. Mengurangi beban kerja manual dosen pembimbing dan koordinator tugas akhir dalam pengecekan kemiripan judul.
2. Memudahkan proses evaluasi dan verifikasi proposal tugas akhir mahasiswa.
3. Menyediakan alat bantu yang efektif dalam proses bimbingan akademik.

1.5 Sistematika Penulisan

Adapun sistematika penulisan dalam penulisan tugas akhir ini terdiri dari beberapa bab dan masing-masing bab membahas dan menguraikan pokok permasalahan yang berbeda, sebagai gambaran disini penulis menyertakan garis-garis besarnya yaitu :

BAB I PENDAHULUAN

Bab ini berisikan latar belakang masalah, rumusan masalah, Batasan masalah, tujuan dan manfaat, serta sistematika penulisan.

BAB II LANDASAN TEORI

Bab ini berisikan tentang tinjauan pustaka yang diambil dari penelitian yang relevan beserta susunan kajian teori yang disesuaikan dengan tema Tugas Akhir.

BAB III METODOLOGI PENELITIAN

Bab ini berisikan tentang tahapan yang dilakukan peneliti dalam mengumpulkan informasi atau data yang dibutuhkan.

BAB II

TINJAUAN PUSTAKA

2.1 Kajian Teori

2.1.1 Cosine Similarity

Dalam bidang text mining, Cosine Similarity merupakan salah satu metode yang paling sering digunakan untuk mengukur tingkat kemiripan antar dua dokumen berdasarkan representasi vektor kata. Setiap dokumen diubah menjadi vektor dalam ruang multidimensi, kemudian tingkat kesamaannya dihitung melalui nilai kosinus dari sudut antara kedua vektor tersebut.

Menurut (Nasrullah, 2024), algoritma Cosine Similarity memiliki keunggulan dalam mendeteksi tingkat kemiripan antar teks dengan tingkat akurasi yang tinggi. Namun, algoritma ini belum mampu mengenali kesamaan makna atau sinonim antar kata yang dibandingkan.

2.1.2 Term Frequency–Inverse Document Frequency (TF-IDF)

Metode TF-IDF merupakan teknik pembobotan yang digunakan untuk menghitung seberapa penting suatu kata terhadap dokumen tertentu. TF-IDF menggabungkan dua konsep utama, yaitu Term Frequency (TF) yang menunjukkan frekuensi kemunculan suatu kata dalam dokumen, dan Inverse Document Frequency (IDF) yang menunjukkan seberapa jarang kata tersebut muncul dalam keseluruhan koleksi dokumen. Semakin tinggi nilai TF-IDF suatu kata, semakin besar tingkat kepentingannya dalam dokumen tersebut.

Menurut (Riyani et al., 2019) metode TF-IDF dikenal efisien, mudah diterapkan, dan memberikan hasil yang akurat untuk perhitungan bobot kata dalam

pengolahan teks. Sementara itu, (Meidelfi et al., 2021) menambahkan bahwa metode ini membantu merepresentasikan setiap kata menjadi nilai numerik untuk kemudian digunakan dalam penghitungan kemiripan antar dokumen.

Tahapan umum dalam penggunaan TF-IDF meliputi:

1. Case Folding:mengubah seluruh huruf menjadi huruf kecil untuk standarisasi teks.
2. Tokenizing:memotong kalimat menjadi kata-kata penyusun (token).
3. Stopword Removal:menghapus kata-kata yang tidak bermakna penting.
4. Stemming:mengubah kata menjadi bentuk dasarnya

2.1.3 Google Spreadsheet

--

2.1.4 Google Appscripts

--

2.2 Penelitian Yang Relevan

Subbab ini membahas beberapa penelitian terdahulu yang berkaitan dengan topik penelitian ini. Uraian tersebut digunakan sebagai acuan dan dasar pengembangan sistem yang diusulkan, serta untuk melihat perbedaan dan keunggulan penelitian yang dilakukan saat ini.

N o	Penulis / Tahun	Topik Penelitian	Metode	Pembahasan	Hasil
1	(Nasrullah, 2024)	Integrasi Tf-Idf Dan Algoritma Cosine Similarity Untuk Deteksi Tingkat Kemiripan	Cosine Similarity dan TF-IDF dengan preprocessing teks lengkap	Mendeteksi kemiripan judul tugas akhir mahasiswa Fakultas Ilmu Komputer UNISAN	Pengujian dengan confusion matrix menghasilkan accuracy rata-rata 89,7%, precision

		Judul Penelitian (Studi Kasus Mahasiswa Fakultas Ilmu Komputer UNISAN Gorontalo)		Gorontalo menggunakan 352 data judul penelitian dari tahun 2019-2022 dengan threshold 0,3 atau 30%	72,4%, dan recall 94,6%. Sistem dapat bekerja dengan baik dalam mengukur tingkat kemiripan judul tugas akhir
2	(Nilawati , Husaini, 2024)	Penggunaan Metode Cosine Similarity dan TF-IDF untuk Klasifikasi Judul Seminar Proposal pada Fakultas Teknik Universitas Jabal Ghafur	Cosine Similarity dan TF-IDF dengan text mining	Membangun sistem klasifikasi judul seminar proposal menggunakan 100 data seminar dengan berbagai kategori bidang ilmu berbasis web menggunakan PHP	Sistem dapat mengklasifikasi kan judul proposal seminar mahasiswa dengan mudah, cepat, dan sesuai dengan kategori yang ditentukan secara otomatis
3	(Eko et al., n.d.)	Studi dan Analisis Algoritma Kemiripan Cosine Similarity untuk Mengidentifikasi Kesamaan Judul Skripsi	Cosine Similarity dengan variasi bobot threshold (0,7; 0,8; 0,9)	Membuat sistem microframework berbasis Python untuk mendeteksi kesamaan judul skripsi di Fakultas Teknologi Informasi Universitas Islam Balitar menggunakan 112 data judul skripsi	Bobot threshold 0,9 menghasilkan akurasi tertinggi sebesar 66% dengan kecepatan akses 0,74 detik. Bobot 0,7 menghasilkan akurasi 38%, dan bobot 0,8 menghasilkan akurasi 43%

BAB III

TINJAUAN PUSTAKA

- 3.1 Tinjauan Umum**
- 3.2 Jenis penelitian**
- 3.3 Desain Penelitian**
- 3.4 Instrumen Penelitian**
- 3.5 dan seterusnya**

DAFTAR PUSTAKA

- Eko, H., Adi, U., Nasiri, A., Informatika, M.T., Framework, M., n.d. STUDI DAN ANALISIS ALGORITMAKEMIRIPAN COSINE SIMILARITY UNTUK MENGIDENTIFIKASIKESAMAAN XII, 1–9.
- Meidelfi, D., Rahmayuni, I., Hidayat, T., Chandra, D., 2021. TF-IDF Implementation for Similarity Checker on The Final Project Title 3, 40–52.
- Nasrullah, A.H., 2024. Integrasi Tf-Idf Dan Algoritma Cosine Similarity Untuk Deteksi Tingkat Kemiripan Judul Penelitian (Studi Kasus Mahasiswa Fakultas Ilmu Komputer UNISAN Gorontalo) 3, 113–118.
- Nilawati, Husaini, J.S., 2024. PENGGUNAAN METODE COSINE SIMILARITY DAN TF-IDF UNTUK KLASIFIKASI JUDUL SEMINAR PROPOSAL PADA 2, 72–79.
- Riyani, A., Zidny, M., Burhanuddin, A., 2019. Penerapan Cosine Similarity dan Pembobotan TF-IDF untuk Mendeteksi Kemiripan Dokumen 2, 23–27.

