

**Penerapan TF-IDF dan *Cosine Similarity* pada Deteksi Kemiripan
Judul Tugas Akhir di 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 TF-IDF dan *Cosine Similarity* pada Deteksi Kemiripan
Judul Tugas Akhir di 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 TF-IDF dan *Cosine Similarity* pada Deteksi Kemiripan Judul Tugas Akhir di STMIK Palangkaraya

Proposal Tugas Akhir Ini Telah Disetujui dan Disahkan

Pembimbing I,

Pembimbing II,

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

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

Mengetahui
Ketua STMIK Palangkaraya,

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

PENGESAHAN

Penerapan *TF-IDF* dan *Cosine Similarity* pada Deteksi Kemiripan Judul Tugas Akhir di STMIK Palangkaraya

Proposal Tugas Akhir ini telah Diseminarkan, Dinalai, 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 PENDAHULUAN	1
1.1 Latar Belakang Masalah.....	1
1.2 Perumusan Masalah	3
1.3 Batasan Masalah.....	3
1.4 Tujuan dan Manfaat	4
1.5 Sistematika Penulisan	5
BAB II TINJAUAN PUSTAKA.....	7
2.1 Kajian Teori	7
2.2 Penelitian Yang Relevan	11
BAB III TINJAUAN PUSTAKA	13
3.1 Tinjauan Umum	13
3.2 Jenis penelitian	14
3.3 Desain Penelitian.....	14
3.4 Instrumen Penelitian.....	16
3.5 dan seterusnya	16
DAFTAR PUSTAKA	18

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 STMIK Palangkaraya, khususnya mahasiswa Program Studi Teknik Informatika. Dalam proses pengajuan judul tugas akhir, sering terjadi permasalahan terkait kemiripan atau kesamaan judul antara mahasiswa dan penelitian sebelumnya. Kemiripan judul dapat menyebabkan duplikasi penelitian, mengurangi orisinalitas karya ilmiah, dan berpotensi menimbulkan tumpang tindih substansi penelitian. Hal ini sejalan dengan pendapat (Purwaningrum et al., 2025) yang menyatakan bahwa tingginya tingkat kemiripan judul tugas akhir dapat menimbulkan berbagai permasalahan, seperti indikasi plagiarisme akademik dan redundansi penelitian yang merugikan dalam konteks pengembangan ilmu pengetahuan.

Berdasarkan data dari website Perpustakaan STMIK Palangkaraya, terdapat [X judul] tugas akhir yang telah terdaftar dalam periode [tahun - tahun] dengan rata-rata [Y mahasiswa] per tahun yang mengajukan judul tugas akhir. Jumlah ini menunjukkan volume data yang cukup besar dan terus meningkat setiap tahunnya, sehingga proses pengecekan manual menjadi semakin kompleks dan memerlukan waktu yang tidak sedikit. Kondisi ini menunjukkan urgensi adanya mekanisme deteksi kemiripan judul yang lebih sistematis, terukur, dan efisien.

Hingga saat ini, proses pengecekan kemiripan judul tugas akhir Program Studi Teknik Informatika STMIK Palangkaraya masih dilakukan secara manual

oleh dosen pembimbing atau koordinator tugas akhir. Belum tersedia sistem terintegrasi untuk pengecekan kesamaan judul. 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 diperiksa juga semakin bertambah. Kondisi ini menyebabkan proses pengecekan manual menjadi tidak efisien dan rentan terhadap kesalahan manusia (*human error*). Hal ini sejalan dengan hasil penelitian yang dilakukan oleh (Salman, 2025), proses pengecekan manual terhadap arsip judul lama sangat tidak efisien, terutama jika tidak tersedia database digital yang terdokumentasi dengan baik. 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 yang dapat mendeteksi kemiripan judul tugas akhir secara otomatis menggunakan algoritma *Term Frequency-Inverse Document Frequency* (TF-IDF) dan *Cosine Similarity*. Menurut (Nasrullah, 2024) menyebutkan bahwa algoritma Cosine Similarity memiliki keunggulan dalam mendeteksi tingkat kemiripan antar teks dengan tingkat akurasi yang tinggi.

Penerapan sistem ini diharapkan dapat meningkatkan efisiensi proses pengecekan judul, mengurangi duplikasi penelitian, dan memudahkan mahasiswa untuk mengakses dokumen tugas akhir terdahulu yang terintegrasi dengan perpustakaan STMIK Palangkaraya.

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 *website* perpustakaan STMIK Palangkaraya?
2. Bagaimana mengimplementasikan algoritma *TF-IDF* dan *Cosine Similarity* dalam mendeteksi tingkat kemiripan judul tugas akhir prodi Teknik Informatika STMIK Palangkaraya?

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 Palangkaraya.
2. Data judul tugas akhir yang digunakan adalah data judul Teknik Informatika yang telah terdaftar dan tersimpan dalam database Perpustakaan STMIK Palangkaraya.
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.

1.4 Tujuan dan Manfaat

Tujuan dari penelitian ini adalah:

1. Menghasilkan sistem berbasis web untuk mendeteksi kemiripan judul tugas akhir yang terintegrasi dengan *website* perpustakaan STMIK Palangkaraya
2. Mengimplementasikan algoritma *TF-IDF* dan *Cosine Similarity* dalam sistem pendeteksi kemiripan judul tugas akhir.

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. Mengintegrasikan database judul tugas akhir Prodi Teknik Informatika pada *website* dengan *scrapping* dari *website* Perpustakaan STMIK Palangkaraya.

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.

Bagi Mahasiswa:

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

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 Tugas Akhir

Berdasarkan Pedoman Akademik STMIK Palangkaraya (2022), tugas akhir merupakan mata kuliah wajib yang harus ditempuh mahasiswa sebagai syarat kelulusan, dengan bentuk dan bobot SKS yang disesuaikan oleh masing-masing program studi. Bagi mahasiswa Diploma, tugas akhir berupa proyek atau paper, sedangkan bagi mahasiswa Program Sarjana berbentuk skripsi.

Menurut Buku Pedoman Tugas Akhir Program Studi Teknik Informatika (2025), pelaksanaan Tugas Akhir memiliki tujuan untuk melatih mahasiswa dalam melakukan kegiatan penelitian secara menyeluruh, yang meliputi pemilihan topik, penyusunan rencana penelitian, pelaksanaan penelitian, evaluasi hasil, serta penulisan dan pertanggungjawaban laporan penelitian secara akademik.

Adapun tujuan penyusunan Tugas Akhir adalah sebagai berikut:

1. Melatih mahasiswa agar mampu memperoleh pengalaman langsung dalam menerapkan dan membandingkan pengetahuan serta keterampilan yang telah diperoleh selama perkuliahan dengan kondisi nyata di masyarakat.
2. Mengembangkan kemampuan mahasiswa dalam menganalisis, mengimplementasikan, dan mengembangkan ilmu pengetahuan serta keterampilan yang dimiliki selama proses penelitian berlangsung.

2.1.2 Data

Secara umum, data dapat diartikan sebagai sekumpulan fakta atau gambaran yang berupa angka maupun simbol lain yang dapat memberikan informasi dan menjadi dasar dalam penarikan kesimpulan. Pengertian data dapat bervariasi tergantung pada bidang yang membahasnya. Dalam bidang komputer, data diartikan sebagai simbol atau sinyal yang digunakan sebagai input, disimpan, serta diproses oleh perangkat komputer untuk menghasilkan output berupa informasi yang berguna.

2.1.3 *Text Mining*

Menurut (Halim, 2024) *Text mining* merupakan suatu proses yang melibatkan penerapan berbagai metode, alat, dan teknik analisis untuk mengekstraksi serta menemukan pengetahuan dari kumpulan dokumen teks dalam jumlah besar. Pengetahuan yang diperoleh melalui proses ini dihasilkan dari pengolahan dan ekstraksi informasi pada sumber data, sehingga menghasilkan kata-kata informatif yang mencerminkan pola serta kecenderungan statistik dalam teks tersebut.

Tahapan umum preprocessing data pada text mining meliputi:

1. *Case Folding*

Tahap ini bertujuan untuk menyeragamkan format teks dengan cara mengubah seluruh huruf menjadi huruf kecil (*lowercase*). Proses ini dilakukan agar sistem tidak membedakan antara huruf besar dan kecil, misalnya kata “Sistem” dan “sistem” dianggap sama.

2. *Tokenizing*

Tokenizing adalah proses pemecahan teks atau kalimat menjadi bagian-bagian yang lebih kecil yang disebut token. Pemisahan bagian yang lebih kecil ini disusun menjadi sebuah *array* yang berfungsi untuk memudahkan proses analisis kata secara individu.

3. *Stopword Removal*

Tahap ini dilakukan untuk menghapus kata-kata yang dianggap tidak memiliki makna penting atau tidak berpengaruh terhadap analisis, seperti kata sambung “dan”, “yang”, “di”, “ke”, dan sebagainya. Tujuannya agar hanya kata-kata yang bermakna informatif yang diproses lebih lanjut. Pada tahap *stopword removal*, penulis menggunakan daftar *stopword* bahasa Indonesia yang diambil dari dataset publik di Kaggle

4. *Stemming*

Stemming merupakan proses mengubah setiap kata ke bentuk dasarnya atau kata dasar (*root word*). Misalnya kata “berlari”, “pelari”, dan “lari-lari” semuanya dikembalikan ke bentuk dasar “lari”. Tahap ini bertujuan untuk mengurangi variasi bentuk kata yang memiliki makna sama, sehingga meminimalkan duplikasi dalam analisis teks.

Pada tahap ini, penulis menggunakan *library Sastrawi.js*, yaitu implementasi algoritma stemming Bahasa Indonesia dalam bahasa *JavaScript*, untuk mengembalikan kata turunan ke bentuk dasarnya secara otomatis.

2.1.4 *Term Frequency–Inverse Document Frequency (TF-IDF)*

Metode TF-IDF merupakan teknik pembobotan yang digunakan untuk menghitung seberapa penting suatu kata pada dokumen terhadap dokumen lainnya. 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.

2.1.5 *Cosine Similarity*

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.

2.1.6 Google Spreadsheet

--

2.1.7 Python

--

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 Judul Penelitian (Studi Kasus Mahasiswa Fakultas Ilmu Komputer UNISAN Gorontalo)	Cosine Similarity dan TF- IDF dengan preprocessi ng teks lengkap	Mendeteksi kemiripan judul tugas akhir mahasiswa Fakultas Ilmu Komputer UNISAN Gorontalo menggunakan 352 data judul penelitian dari tahun 2019- 2022 dengan threshold 0,3 atau 30%	Pengujian dengan confusion matrix menghasilkan accuracy rata- rata 89,7%, precision 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., 2023)	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 mendekripsi 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

Penelitian ini difokuskan pada penerapan metode *TF-IDF (Term Frequency–Inverse Document Frequency)* dan *Cosine Similarity* untuk mendeteksi tingkat kemiripan judul Tugas Akhir. Tujuan dari penelitian ini adalah untuk membantu proses verifikasi dan pengajuan judul Tugas Akhir mahasiswa dengan cara membandingkan judul yang akan diajukan dengan kumpulan judul yang telah terdaftar sebelumnya di perpustakaan STMIK Palangkaraya.

Melalui sistem ini, diharapkan dapat memberikan kemudahan bagi dosen pembimbing Tugas Akhir dalam memberikan bimbingan dan saran kepada mahasiswa, serta membantu menghindari duplikasi atau kemiripan judul penelitian di lingkungan Program Studi Teknik Informatika STMIK Palangkaraya.

Selain itu, sistem ini juga bermanfaat bagi mahasiswa, karena mereka dapat mengecek sendiri tingkat kemiripan judul yang akan diajukan sebelum proses pengajuan resmi dilakukan. Dengan demikian, mahasiswa dapat menjadikan sistem ini sebagai referensi awal dalam menentukan judul yang unik, relevan, dan belum pernah digunakan oleh mahasiswa lain sebelumnya.

Objek penelitian ini adalah judul-judul Tugas Akhir mahasiswa Program Studi Teknik Informatika STMIK Palangkaraya yang telah terarsip di *website* Perpustakaan STMIK Palangkaraya. Data tersebut digunakan sebagai acuan dalam

proses pengujian sistem pendekripsi kemiripan judul berbasis metode *TF-IDF* dan *Cosine Similarity*.

3.2 Jenis penelitian

Jenis penelitian yang digunakan dalam tugas akhir ini adalah penelitian pengembangan (*Research and Development*). Penelitian ini berfokus pada proses perancangan dan pembangunan sebuah sistem yang dapat mendekripsi tingkat kemiripan judul Tugas Akhir menggunakan metode *TF-IDF* (*Term Frequency–Inverse Document Frequency*) dan *Cosine Similarity*.

Hasil dari penelitian ini berupa produk aplikasi berbasis web yang dirancang untuk membantu mahasiswa dan dosen pembimbing dalam mengecek kesamaan judul Tugas Akhir secara otomatis. Dengan adanya sistem ini, diharapkan proses pengajuan judul menjadi lebih mudah, efisien, serta mampu mengurangi kemungkinan terjadinya duplikasi judul penelitian di lingkungan Program Studi Teknik Informatika STMIK Palangkaraya.

3.3 Desain Penelitian

Desain penelitian yang digunakan dalam pengembangan sistem pendekripsi kemiripan judul tugas akhir ini adalah penelitian dan pengembangan (*Research and Development*) dengan model yang disederhanakan menjadi lima tahapan utama. Model ini dipilih karena penelitian tidak hanya berfokus pada pembuatan sistem, tetapi juga pada pengujian dan penyempurnaan agar sistem dapat memberikan hasil yang akurat serta mudah digunakan oleh pengguna.

Adapun tahapan penelitian yang dilakukan adalah sebagai berikut:

1. Studi Literatur dan Pengumpulan Data

Tahap ini bertujuan untuk memperoleh landasan teori dan pemahaman mendalam mengenai metode *TF-IDF* dan *Cosine Similarity* yang digunakan dalam mendeteksi kemiripan teks. Selain itu, dilakukan pula pengumpulan data berupa judul-judul tugas akhir dari arsip atau repositori STMIK Palangkaraya yang akan dijadikan *dataset* pengujian. Hasil dari tahap ini berupa kumpulan referensi teori dan dataset awal yang siap digunakan dalam tahap perancangan sistem.

2. Perencanaan dan Perancangan Sistem

Pada tahap ini dilakukan identifikasi kebutuhan sistem. dirancang pula alur kerja sistem, struktur basis data, serta rancangan antarmuka pengguna (*user interface*) yang sederhana dan mudah dipahami. Perencanaan juga mencakup penentuan kriteria kemiripan judul dan rancangan mekanisme perhitungan *TF-IDF* serta *Cosine Similarity*.

3. Pembuatan Prototipe dan Implementasi Awal

Tahap ini merupakan proses pembangunan sistem berbasis web sesuai rancangan yang telah dibuat sebelumnya. Sistem dikembangkan menggunakan metode *TF-IDF* dan *Cosine Similarity* untuk menghitung tingkat kemiripan antarjudul. Fitur utama yang dikembangkan meliputi proses input judul, *preprocessing* teks (*case folding*, *tokenizing*, *stopword removal*, dan *stemming*), perhitungan nilai kemiripan, serta tampilan hasil perbandingan dalam bentuk skor kemiripan.

4. Pengujian dan Evaluasi Sistem

Setelah sistem selesai diimplementasikan, dilakukan pengujian untuk memastikan sistem berfungsi sesuai dengan kebutuhan. Pengujian dilakukan dalam dua tahap, yaitu:

- a. Pengujian fungsional, untuk memastikan seluruh fitur berjalan dengan baik dan hasil perhitungan sesuai dengan logika metode yang digunakan.
- b. Pengujian validitas deteksi, dengan menggunakan dataset berlabel untuk mengevaluasi tingkat akurasi sistem berdasarkan metrik precision, recall, dan F1-score. Selain itu, dilakukan pula uji coba pengguna (User Acceptance Test) yang melibatkan dosen dan mahasiswa untuk menilai kemudahan penggunaan dan relevansi hasil yang ditampilkan.

5. Revisi dan Implementasi Akhir

Berdasarkan hasil pengujian dan umpan balik dari pengguna, dilakukan perbaikan terhadap sistem baik dari sisi tampilan, performa, maupun keakuratan hasil deteksi. Tahap ini menghasilkan versi akhir aplikasi web pendekripsi kemiripan judul tugas akhir yang siap digunakan dalam proses pengajuan judul di lingkungan Program Studi Teknik Informatika STMIK Palangkaraya.

3.4 Instrumen Penelitian

3.5 dan seterusnya

DAFTAR PUSTAKA

- Eko, H., Adi, U., Nasiri, A., Informatika, M.T., Framework, M., 2023. STUDI DAN ANALISIS ALGORITMA KEMIRIPAN COSINE SIMILARITY UNTUK MENIDENTIFIKASI KESAMAAN JUDUL SKRIPSI. PROSIDING SEMINAR ILMIAH SISTEM INFORMASI DAN TEKNOLOGI INFORMASI 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.
- Purwaningrum, S., Susanto, A., Prabowo, A.S., 2025. Kombinasi Algoritma TF-IDF dan Weighted Dice Similarity Untuk Pengukuran Kemiripan Judul Tugas Akhir 16, 328–337. <https://doi.org/10.35970/infotekmesin.v16i2.2812>
- Riyani, A., Zidny, M., Burhanuddin, A., 2019. Penerapan Cosine Similarity dan Pembobotan TF-IDF untuk Mendeteksi Kemiripan Dokumen 2, 23–27.
- Salman, 2025. SISTEM PENDETEKSI KEMIRIPAN JUDUL SKRIPSI BERBASIS WEB 5. <https://doi.org/10.59818/jpi.v5i5.1964>

