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
2026

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
2026

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	Ir. Abdul Hadi, S.T., M.Kom.
NIK. 197502232009002	NIK. 198505102021101

Mengetahui	Mengesahkan
Ketua STMIK Palangkaraya,	Ketua Prodi Teknik Informatika,
Ir. Hj. Siti Maryamah, MM.	Ir. Abdul Hadi, S.T., M.Kom.
NIK. 196902021995004	NIK. 198505102021101

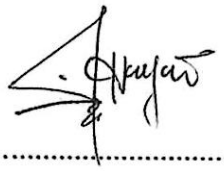
PENGESAHAN

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

Proposal Tugas Akhir ini telah Diseminarkan, Dinilai, dan Disahkan
Oleh Tim Seminar pada Tanggal 13 Januari 2026

Tim Seminar Proposal :

1. Sulistyowati, S.Kom., M.Cs. ..... 
   Ketua

2. Dr. Maura Widyaningsih, S.Kom,M.Cs ..... 
   Sekretaris

3. Ir. Abdul Hadi, S.T., M.Kom. ..... 
   Anggota

DAFTAR ISI

PERSETUJUAN ..... **i**
PENGESAHAN ..... **ii**
DAFTAR TABEL ..... **iv**
DAFTAR GAMBAR ..... **v**
DAFTAR LAMPIRAN ..... **i**
BAB I PENDAHULUAN **1**
1.1 Latar Belakang Masalah 1
1.2 Perumusan Masalah 3
1.3 Batasan Masalah 4
1.4 Tujuan dan Manfaat 4
1.5 Sistematika Penulisan 6
BAB II TINJAUAN PUSTAKA **7**
2.1 Kajian Teori 7
2.2 Penelitian Yang Relevan 20
BAB III METODOLOGI PENELITIAN **24**
3.1. Tinjauan Umum 24
3.2. Jenis penelitian 25
3.3. Desain Penelitian 25
3.4. Instrumen Penelitian 27
3.5. Analisis Kebutuhan 28
DAFTAR PUSTAKA

DAFTAR TABEL

Tabel 2.1 Contoh Proses *Case Folding* 9
Tabel 2.2 Contoh Proses *Tokenizing* 9
Tabel 2.3 Contoh Proses *Stopword Removal* 10
Tabel 2.4 Contoh Proses *Stemming* 10
Tabel 2.5 Penelitian Relevan 20
Tabel 3.1 Perangkat Keras 29
Tabel 3.2 Perangkat Lunak 29
Tabel 3.3 Kategori Tingkat Kemiripan Judul 30
Tabel 3.4 Hasil *Case Folding* pada Judul Tugas Akhir 31
Tabel 3.5 Hasil *Tokenizing* pada Judul Tugas Akhir 31
Tabel 3.6 Hasil *Stopword Removal* pada Judul Tugas Akhir 32
Tabel 3.7 Hasil *Stemming* pada Judul Tugas Akhir 33
Tabel 3.8 Frekuensi Kemunculan *Term* (TF) pada Setiap Dokumen 33
Tabel 3.9 Perhitungan Nilai IDF untuk Setiap Term 34
Tabel 3.10 Nilai Bobot TF-IDF Setiap *Term* pada Masing-Masing Dokumen 35
Tabel 3.11 Hasil Akhir *Cosine Similarity* 37
Tabel 3.12 Database Spreadsheets 48
Tabel 3.13 Jadwal Penelitian 49

DAFTAR GAMBAR

Gambar 2.1 Metode *Waterfall* 15
Gambar 3.1 *Flowchart* Alur Kerja *Website* 38
Gambar 3.2 *Flowchart* Alur Kerja *Scraping* 40
Gambar 3.3 *Usecase Diagram* proses 41
Gambar 3.4 *Activity Diagram* Alur Sistem 42
Gambar 3.5 Desain Halaman Utama 43
Gambar 3.6 Desain Halaman Hasil 44
Gambar 3.7 Desain Halaman Detail 45
Gambar 3.8 Desain Halaman Daftar Judul 46

DAFTAR LAMPIRAN

Lampiran 1. Surat tugas pembimbing tugas akhir
Lampiran 2. Lembar konsultasi bimbingan Tugas Akhir
Lampiran 3. Surat izin penelitian
Lampiran 4. Surat tugas penguji seminar
Lampiran 5. Daftar Hadir Peserta Seminar

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Tugas akhir merupakan salah satu syarat kelulusan bagi mahasiswa STMIK Palangkaraya. Dalam proses pengajuan judul, sering ditemui permasalahan berupa kemiripan atau kesamaan judul antara mahasiswa dengan penelitian terdahulu. Kemiripan judul tidak hanya mengurangi orisinalitas karya ilmiah, tetapi juga berpotensi menimbulkan duplikasi topik serta tumpang tindih substansi penelitian. Hal ini sejalan dengan pendapat (Purwaningrum et al., 2025) yang menegaskan 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, tercatat sebanyak 2174 judul tugas akhir yang tersimpan dalam repositori hingga tahun 2024. Jumlah tersebut menunjukkan bahwa volume arsip penelitian yang dikelola cukup besar dan terus bertambah seiring dengan meningkatnya jumlah mahasiswa. Peningkatan jumlah data tersebut secara langsung berdampak pada bertambahnya beban kerja dosen pembimbing dan koordinator tugas akhir dalam proses verifikasi judul. Pada praktiknya, proses pemeriksaan kemiripan judul masih dilakukan secara manual dengan cara mencocokkan judul baru terhadap daftar judul lama satu per satu, sehingga memerlukan waktu yang lama dan berpotensi menimbulkan ketidaktepatan dalam pengambilan keputusan.

Meskipun *website* resmi perpustakaan kampus telah menyediakan arsip Tugas Akhir secara *online*, sistem tersebut belum dilengkapi dengan fitur pencarian kemiripan judul secara otomatis. Akses data hanya berupa tampilan informasi judul dan dokumen, tanpa kemampuan analisis berbasis teks yang dapat mendukung proses verifikasi. Penelitian oleh (Salman, 2025), menegaskan bahwa digitalisasi arsip tanpa disertai mekanisme pencarian dan analisis otomatis masih belum cukup untuk meningkatkan efisiensi kerja, karena proses peninjauan tetap harus dilakukan secara manual oleh dosen atau pengelola repositori. Ketiadaan fitur analisis otomatis inilah yang membuat proses pengecekan judul di STMIK Palangkaraya rentan mengalami keterlambatan dan ketidakefektifan.

Selain itu, *website* Perpustakaan STMIK Palangkaraya tidak menyediakan *API* resmi atau *endpoint* integrasi untuk mengakses data secara otomatis, sehingga pembaruan data harus dilakukan dengan cara *web scraping* secara periodik. Kondisi ini menimbulkan kebutuhan akan sistem yang mampu mengotomatisasi pengambilan data, membersihkan data (*preprocessing*), serta melakukan perhitungan kemiripan judul menggunakan algoritma khusus. Dengan demikian, sistem pendeteksi kemiripan judul tidak bergantung pada input manual dan mampu memperbarui database secara berkala sesuai dengan perkembangan arsip penelitian.

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) algoritma *Cosine Similarity* memiliki keunggulan dalam mengukur kedekatan antar teks karena mampu menghitung derajat kemiripan

berdasarkan representasi vektor, sehingga tepat digunakan dalam proses pencarian dan deteksi kemiripan dokumen. Sementara itu, TF-IDF efektif dalam memberikan bobot pada kata-kata penting sehingga sistem dapat mengenali istilah dominan dalam judul penelitian. Kombinasi kedua algoritma ini terbukti banyak digunakan dalam sistem rekomendasi, pencarian dokumen, dan deteksi plagiarisme.

Dengan adanya sistem deteksi kemiripan judul yang terintegrasi dengan repositori digital kampus, proses verifikasi judul dapat dilakukan dengan lebih cepat, akurat, dan efisien. Sistem ini tidak hanya membantu dosen dalam mengurangi beban pemeriksaan manual, tetapi juga memudahkan mahasiswa untuk mengetahui tingkat kesesuaian atau potensi kemiripan judul yang mereka ajukan. Selain itu, penggunaan teknologi seperti *web scraping, preprocessing* teks otomatis, serta penyimpanan data yang terpusat pada platform digital dengan *Google Sheets* yang memungkinkan sistem tetap ringan, mudah dipelihara, dan dapat diperbarui tanpa proses input data satu per satu. Oleh karena itu, pengembangan sistem pendeteksi kemiripan judul berbasis *web* menjadi solusi yang relevan dan diperlukan dalam mendukung manajemen tugas akhir di STMIK Palangkaraya.

1.2 Perumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, maka rumusan masalah dalam penelitian ini adalah bagaimana merancang dan membangun sistem berbasis web yang terintegrasi dengan *website* perpustakaan STMIK Palangkaraya untuk mendeteksi kemiripan judul tugas akhir, serta bagaimana mengimplementasikan

algoritma TF-IDF dan *Cosine Similarity* dalam proses pendeteksian tingkat kemiripan judul tugas akhir pada 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 tugas akhir.
2. Data judul tugas akhir yang digunakan adalah data judul yang telah terdaftar dan tersimpan di *website* 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.
5. Pembaruan data dilakukan melalui proses *web scraping* yang dijalankan oleh Koordinator 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.
3. Menghasilkan sistem *web Scraping* untuk pengambilan data judul.

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 dengan *website* Perpustakaan STMIK Palangkaraya melalui mekanisme *web scraping*.

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. Memberikan kemandirian kepada mahasiswa dalam melakukan pengecekan kemiripan judul tugas akhir sebelum pengajuan judul.
2. Membantu mahasiswa mencegah duplikasi judul penelitian sejak tahap awal secara mandiri.
3. Meningkatkan efisiensi proses pengajuan proposal tugas akhir dengan meminimalkan revisi akibat kesamaan judul.

4. Mempermudah mahasiswa dalam menemukan judul-judul penelitian terdahulu sebagai bahan referensi dan pengembangan ide penelitian.

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.

2.1.2 *Website*

Website adalah suatu aplikasi yang diakses menggunakan *web browser* melalui suatu jaringan seperti internet atau intranet. *Website* dapat diartikan suatu kumpulan-kumpulan halaman yang menampilkan berbagai macam informasi teks, data, gambar, animasi, suara, video maupun gabungan dari semuanya, baik statis maupun dinamis, yang dimana membentuk satu rangkaian bangunan yang saling berkaitan dan masing-masing dihubungkan dengan jaringan halaman atau hyperlink (Amanah Amnun Zulfa, Tatang Ibrahim, 2025).

Dalam penelitian ini, *website* berperan sebagai platform utama untuk menampilkan hasil pengecekan judul serta menyediakan antarmuka bagi pengguna, untuk mahasiswa, dalam melakukan input judul tugas akhir.

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.

Sebelum proses perhitungan dapat dilakukan, data teks perlu melalui tahap *preprocessing* untuk memastikan bahwa teks berada dalam bentuk yang bersih, terstruktur, dan siap diproses oleh algoritma. Tahapan ini penting agar hasil analisis lebih akurat terhadap judul. Adapun tahapan umum *preprocessing* dalam *text mining* meliputi:

a. *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 huruf kecil. Contoh proses case folding ditunjukkan pada Tabel 2.1.

Tabel 2.1 Contoh Proses *Case Folding*

Teks Asli	Hasil *Case Folding*
Sistem Informasi SANGAT Penting	sistem informasi sangat penting
Sekolah Tinggi Manajemen Informatika dan Komputer	sekolah tinggi manajemen informatika dan komputer
ANALISIS Data Mining	analisis data mining

b. *Tokenizing*

Tokenizing merupakan proses pemecahan teks atau kalimat menjadi bagian-bagian yang lebih kecil yang disebut *token*. Proses ini bertujuan untuk memudahkan analisis setiap kata secara individual dengan menyusunnya ke dalam bentuk *array*. Contoh hasil proses *tokenizing* ditunjukkan pada Tabel 2.2.

Tabel 2.2 Contoh Proses *Tokenizing*

Teks Input	Hasil *Tokenizing*
sistem informasi sangat penting	["sistem", "informasi", "sangat", "penting"]
sekolah tinggi manajemen informatika dan komputer	["sekolah", "tinggi", "manajemen", "informatika", "dan", "komputer"]
analisis data mining	["analisis", "data", "mining"]

c. *Stopword Removal*

Tahap ini dilakukan untuk menghapus kata-kata yang dianggap tidak memiliki makna penting atau tidak berpengaruh terhadap proses analisis. Tujuannya adalah agar hanya kata-kata yang bersifat informatif yang diproses lebih lanjut. Pada penelitian ini, proses *stopword removal* menggunakan daftar *stopword* bahasa Indonesia yang diperoleh dari

dataset publik Kaggle. Contoh hasil proses *stopword removal* ditunjukkan pada Tabel 2.3.

Tabel 2.3 Contoh Proses *Stopword Removal*

Kata Stopword Removal	Kata Setelah *Stopword Removal*
["sistem", "informasi", "sangat", "penting"]	["sistem", "informasi", "penting"]
["sekolah", "tinggi", "manajemen", "informatika", "dan", "komputer"]	["sekolah", "tinggi", "manajemen", "informatika", "komputer"]
["analisis", "data", "mining"]	["analisis", "data", "mining"]

d. *Stemming*

Stemming merupakan proses mengubah setiap kata ke bentuk dasarnya. Tahap ini bertujuan untuk mengurangi variasi bentuk kata yang memiliki makna sama sehingga meminimalkan duplikasi dalam analisis teks. Pada penelitian ini, proses *stemming* dilakukan menggunakan *library Sastrawi.js*, yaitu implementasi algoritma *stemming* Bahasa Indonesia berbasis *JavaScript*. Contoh hasil proses *stemming* ditunjukkan pada Tabel 2.4.

Tabel 2.4 Contoh Proses *Stemming*

Kata Sebelum Stemming	Kata Setelah *Stemming*
Berlari	lari
Pembelajaan	ajar
Menganalisis	analisis
Menggunakan	guna
Pendukung	dukung

2.1.4 *Web scraping*

Web scraping adalah teknik yang digunakan untuk mengekstraksi data dari suatu *website* secara otomatis. Teknik ini melibatkan penargetan halaman *web* tertentu, mengambil kode HTML yang mendasarinya, memparsing data yang

relevan, dan menyimpannya ke dalam sistem file atau basis data untuk dianalisis lebih lanjut (Dikilitaş et al., 2024). Dalam penelitian ini, *web scraping* digunakan untuk mengambil data judul tugas akhir dari *website* perpustakaan sebagai sumber data. Proses *scraping* dilakukan dengan memanfaatkan klasifikasi dokumen "TA TI" sebagai acuan seleksi data yang akan diambil.

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

Metode *Term Frequency–Inverse Document Frequency* (TF-IDF) merupakan teknik pembobotan kata yang banyak digunakan dalam pengolahan teks untuk menentukan tingkat kepentingan suatu kata dalam sebuah dokumen terhadap kumpulan dokumen lainnya. TF-IDF bekerja dengan menggabungkan dua konsep utama, yaitu *Term Frequency* (TF) dan *Inverse Document Frequency* (IDF).

Term Frequency (TF) menggambarkan seberapa sering suatu kata muncul dalam satu dokumen. Semakin sering kata tersebut muncul, semakin tinggi nilai TF-nya. Nilai TF dapat dihitung menggunakan rumus berikut:

$$TF\ (t,d) = \ f(t,d) \qquad (2.1)$$

Keterangan:

TF = term (kata) yang dicari

d = dokumen

TF(t,d)= frekuensi kemunculan term t dalam dokumen d

F(t,d) = jumlah kemunculan term *t* dalam dokumen *d*.

Sementara itu, Inverse *Document Frequency* (IDF) digunakan untuk mengukur tingkat kelangkaan suatu kata pada seluruh kumpulan dokumen. Kata yang jarang muncul akan memiliki nilai IDF yang tinggi, sedangkan kata yang sering muncul pada banyak dokumen akan memiliki nilai IDF yang rendah. Rumus IDF dituliskan sebagai berikut:

$$idf(t) = \log\left(\frac{N}{df_t}\right) + 1 \qquad (2.2)$$

Keterangan:

N = Total jumlah dokumen dalam korpus (*collection*)

Dft = Document Frequency, yaitu jumlah dokumen yang mengandung *term t*

Nilai akhir TF-IDF diperoleh dengan mengalikan nilai TF dan IDF, sehingga menunjukkan tingkat kepentingan suatu kata pada dokumen tertentu sekaligus mempertimbangkan kelangkaannya pada dokumen lain. Perhitungan TF-IDF dirumuskan sebagai berikut:

$$TF - IDF(t,d) = TF(t,d) \times IDF(t) \qquad (2.3)$$

Menurut (Riyani et al., 2019), metode TF-IDF dikenal sebagai teknik pembobotan yang efisien, mudah diterapkan, dan akurat dalam menentukan bobot kata pada proses pengolahan teks. Sementara itu, (Meidelfi et al., 2021) menambahkan bahwa TF-IDF mampu mengubah setiap kata menjadi representasi numerik sehingga dapat digunakan dalam perhitungan kemiripan antar dokumen, termasuk dalam proses deteksi kemiripan judul tugas akhir.

2.1.6 *Cosine Similarity*

Cosine Similarity merupakan salah satu metode yang paling banyak digunakan untuk mengukur tingkat kemiripan antara dua dokumen berdasarkan representasi vektornya. Dalam pendekatan ini, setiap dokumen terlebih dahulu diubah menjadi sebuah vektor numerik, umumnya menggunakan bobot kata seperti TF-IDF. Setelah masing-masing dokumen direpresentasikan dalam bentuk vektor, nilai kemiripan dihitung berdasarkan nilai kosinus dari sudut yang terbentuk antara kedua vektor tersebut di dalam ruang multidimensi.

Metode ini bekerja dengan prinsip bahwa semakin kecil sudut antara dua vektor, semakin besar nilai kosinusnya, dan semakin mirip pula dokumen tersebut. Sebaliknya, jika sudut antara vektor semakin besar, nilai kosinus akan semakin kecil, sehingga tingkat kemiripannya juga menurun. Pendekatan ini efektif karena tidak dipengaruhi oleh panjang dokumen, tetapi lebih menekankan pada pola distribusi kata.

Secara matematis, *Cosine Similarity* antara dua vektor *A* dan *B* dapat dihitung dengan rumus:

$$Cosine\ Similarity\ (A, B) = \frac{A \cdot B}{||A|| \times ||B||} \qquad (2.4)$$

Keterangan:

A,B = Vektor representasi dokumen (hasil TF-IDF)

A·B = Dot product (perkalian titik) antara vektor A dan B

||*A*|| = Magnitude (norma Euclidean) vektor A

||*B*|| = Magnitude (norma Euclidean) vektor B

Rumus tersebut dapat dijelaskan lebih lanjut sebagai berikut:

$$A \cdot B = \sum_{i=1}^{n} A_i \times B_i \qquad (2.5)$$

$$||A|| = \sqrt{\sum_{i=1}^{n} A_i^2} \quad \text{dan} \quad ||B|| = \sqrt{\sum_{i=1}^{n} B_i^2} \qquad (2.6)$$

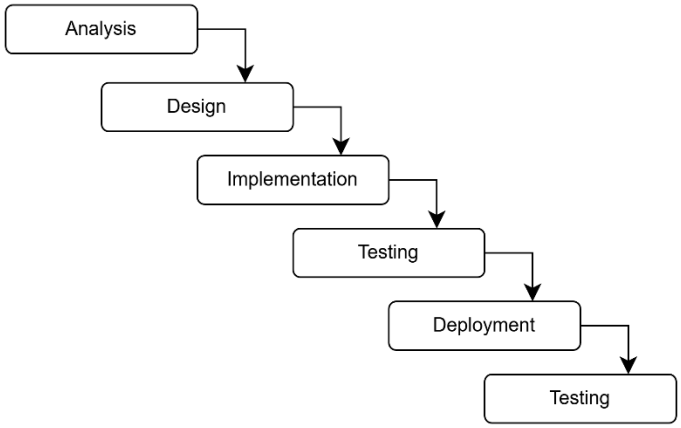
Hasil perhitungan *Cosine Similarity* berada pada rentang 0 hingga 1, di mana:

a. nilai 1 menunjukkan bahwa dua dokumen sangat mirip (arah vektornya sama),
b. nilai 0 menunjukkan tidak ada kemiripan (vektor saling tegak lurus).

Menurut (Nasrullah, 2024), algoritma *Cosine Similarity* memiliki keunggulan dalam mendeteksi tingkat kemiripan antar teks dengan tingkat akurasi yang tinggi.

2.1.7 *Waterfall*

Menurut (Fitria Nur Hasanah M.Pd, 2020) Model air terjun (*Waterfall Model*) adalah pendekatan klasik dalam pengembangan perangkat lunak yang menggambarkan metode pengembangan linier dan berurutan. Ini terdiri dari lima hingga tujuh fase, setiap fase didefinisikan oleh tugas dan tujuan yang berbeda, di mana keseluruhan fase menggambarkan siklus hidup perangkat lunak hingga pengirimannya. Setelah fase selesai, langkah pengembangan selanjutnya mengikuti dan hasil dari fase sebelumnya mengalir ke fase berikutnya.



Gambar 2.1 Metode *Waterfall*

(Sumber: Fitria Nur Hasanah M.Pd, 2020)

Dikutip dari (Fitria Nur Hasanah M.Pd, 2020) ada enam tahapan dalam metode *Waterfall* sebagai berikut

a. Analisis

   Pada tahap pertama dilakukan proses pengumpulkan kebutuhan secara lengkap kemudian dianalisis dan didefinisikan kebutuhan yang harus dipenuhi oleh program yang akan dibangun. Fase ini harus dikerjakan secara lengkap untuk bisa menghasilkan desain yang lengkap.

b. Desain

   Pada tahap ini pengembang akan menghasilkan sebuah sistem secara keseluruhan dan menentukan alur perangkat lunak hingga algoritma yang detail.

c. Implementasi

   Implementasi adalah tahapan di mana seluruh desain diubah menjadi kode program. Kode program yang dihasilkan masih berupa modul-modul yang akan diintegrasikan menjadi sistem yang lengkap.

d. Pengujian

Tahap ini dilakukan dengan menggabungkan modul-modul yang sudah dibuat dan melakukan pengujian untuk mengetahui apakah *software* yang dibuat telah sesuai dengan desainnya serta untuk memastikan apakah terdapat kesalahan pada fungsi *software* atau tidak.

e. Verifikasi

Verifikasi adalah klien atau pengguna menguji apakah sistem tersebut telah sesuai dengan yang disetujui.

f. Pemeliharaan

Pada tahap ini dilakukan proses instalasi dan perbaikan sistem sesuai yang disetujui

2.1.8 Perangkat Lunak dan Alat yang digunakan

a. Figma

Figma adalah salah satu *design tool* yang digunakan untuk membuat tampilan aplikasi *mobile*, *desktop*, *website*, dan berbagai kebutuhan desain antarmuka lainnya. Figma dapat digunakan pada sistem operasi Windows, Linux, maupun macOS selama terhubung ke internet. Secara umum, Figma banyak digunakan oleh profesional di bidang *UI/UX*, *web design*, dan bidang kreatif sejenis (Muhyidin et al., 2020).

b. Draw.io

Menurut (Hendrawan et al., 2023) Draw.io adalah perangkat lunak yang digunakan untuk membuat dan mensimulasikan rancangan jaringan

komputer. Perangkat lunak ini menawarkan kemampuan simulasi, penyusunan tata letak (*layout*), serta kolaborasi sehingga memudahkan pembuatan konsep teknologi yang kompleks. Selain itu, Draw.io menyediakan berbagai komponen yang umum digunakan dalam struktur jaringan, sehingga pengguna dapat dengan mudah membuat simulasi jaringan untuk melihat cara kerja perangkat serta proses pengiriman data dari satu komputer ke komputer lainnya.

c. *Javascript*

JavaScript merupakan bahasa pemrograman yang digunakan dalam pengembangan web, di mana sebagian besar situs modern memanfaatkan JavaScript untuk menghadirkan interaktivitas pada halaman web. Hampir semua browser modern pada perangkat desktop, tablet, maupun ponsel telah menyertakan dukungan penuh terhadap JavaScript, menjadikannya salah satu bahasa pemrograman yang paling umum dan luas digunakan. JavaScript dikenal sebagai bahasa yang dinamis serta memiliki mekanisme interpretasi yang mendukung paradigma pemrograman berorientasi objek maupun pemrograman fungsional (Fandopa and Santoso, 2022).

d. Google Spreadsheets

Google Sheets merupakan aplikasi spreadsheet berbasis *cloud* yang dikembangkan oleh Google sebagai bagian dari Google Workspace. Aplikasi ini menyediakan struktur baris, kolom, dan cell untuk mengolah, menyimpan,

serta menganalisis data secara terorganisir. Sebagai spreadsheet berbasis *web*, Google Sheets memungkinkan pengguna melakukan akses dan kolaborasi *real-time*, sehingga setiap perubahan dapat dilihat oleh seluruh pengguna yang memiliki izin akses. Google Sheets Memungkinkan, integrasi otomatis, kolaborasi daring, dan kemampuan impor–ekspor menjadikan Google Sheets fleksibel untuk berbagai kebutuhan pengelolaan data (Google, 2024).

Google Sheets digunakan sebagai basis data sederhana dalam aplikasi *web*. Dalam penelitian ini, Google Sheets dimanfaatkan sebagai media penyimpanan data judul tugas akhir sebelum diproses melalui tahapan text mining menggunakan metode TF-IDF dan Cosine Similarity.

e. Python

Python adalah bahasa yang cocok untuk belajar dan sebuah pemrograman yang nyata. Python sangat kuat dan sebuah bahasa pemrograman tingkat tinggi yang dibuat oleh seorang *programmer* bernama Guido van Rossum (Halim, 2024). Proses scraping dilakukan menggunakan bahasa pemrograman Python karena memiliki berbagai *library* pendukung seperti *Requests* dan *BeautifulSoup* yang memudahkan dalam mengambil dan mem-parsing elemen-elemen pada halaman *web* perpustakaan. Dengan demikian, Python berperan sebagai perangkat utama dalam tahap pengumpulan data pada penelitian ini.

f. Google Drive

Google Drive adalah media penyimpanan data secara daring (*online*) berbasis internet (*cloud*) yang dikembangkan oleh Google dan diperkenalkan secara resmi pada tanggal 24 April 2012. Pada dasarnya layanan Google Drive sama seperti *cloud storage* lain seperti DropBox atau OneDrive (Suwarya, 2021).

(Trilaksono et al., 2020) mengidentifikasi lima manfaat utama Google Drive yaitu kemampuan mengirim berbagai jenis file, sinkronisasi otomatis pada komputer, laptop dan smartphone, integrasi dengan aplikasi Google untuk kolaborasi *real-time*, pengelolaan hak akses yang fleksibel dengan pengaturan private atau public, serta penyediaan ruang penyimpanan yang besar. Dalam konteks pendidikan, (Kurniawan et al., 2023) menambahkan bahwa Google Drive menawarkan aksesibilitas untuk mengakses data kapan saja dan dimana saja, skalabilitas untuk perluasan kapasitas penyimpanan tanpa perangkat tambahan, serta keamanan data yang terjamin bahkan saat terjadi kerusakan perangkat keras. Google Drive juga efektif sebagai solusi backup data dalam lingkungan pendidikan, memungkinkan siswa menyimpan tugas sekolah dengan aman dan berbagi file dengan guru secara terstruktur.

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.

Tabel 2.5 Penelitian Relevan

No	Penulis / Tahun	Topik Penelitian	Metode	Pembahasan	Hasil
1	(Nasrullah, 2024)	Integrasi TF-IDF dan *Cosine Similarity* untuk Deteksi Tingkat Kemiripan Judul Penelitian	TF-IDF, *Cosine Similarity, Text Preprocessing, Confusion Matrix.*	Mengembangkan aplikasi deteksi plagiarisme dokumen mahasiswa berbasis *web* dengan membandingkan dokumen pengguna terhadap database statis yang diinput secara manual.	Hasil perhitungan manual sebesar 4% dan hasil aplikasi 4,34% menunjukkan konsistensi perhitungan yang baik.
2	(Nilawati, Husaini, 2024)	Penggunaan Metode *Cosine Similarity* dan *TF-IDF* untuk Klasifikasi Judul Seminar Proposal	TF-IDF, *Cosine Similarity, Text Mining, PHP, MySQL*	Mengembangkan sistem klasifikasi judul seminar proposal untuk membantu pengelompokan judul skripsi. Proses text mining dilakukan mulai dari tahap preprocessing	Sistem mampu melakukan klasifikasi judul secara otomatis dan mempercepat proses pengelompokan data.

				hingga perhitungan tingkat kemiripan dokumen, sehingga sistem mampu melakukan klasifikasi judul secara otomatis dan efisien menggunakan data latih berupa data dummy.	
3	(Eko et al., 2023)	Cosine Similarity sebagai metode pendeteksi kemiripan judul dengan python	Cosine Similarity dan TF-IDF	Fokus pembahasan bertumpu pada permasalahan belum tersedianya sistem validasi judul skripsi. Sistem yang dikembangkan melakukan preprocessing teks, menghitung vektor kata, lalu menampilkan persentase kemiripan antar judul. Pendekatan ini mempermudah dosen dalam memverifikasi apakah judul	Sistem yang dihasilkan mampu menampilkan persentase kemiripan judul skripsi dan berfungsi sebagai alat bantu validasi pengajuan judul. Sistem ini membantu pembimbing dalam mempercepat proses pengecekan kesamaan judul yang diajukan mahasiswa.

				tersebut sudah pernah diajukan.	
4	(Meidelfi et al., 2021)	Deteksi kemiripan judul tugas akhir mahasiswa TRPL PNP	*Cosine Similarity, TF-IDF, Preprocessing (Case Folding, Tokenizing, Filtering, Stemming)*	Sistem dibangun untuk membantu KBK memeriksa kesamaan judul tugas akhir yang diajukan mahasiswa dengan judul yang sudah ada. Proses meliputi preprocessing teks, pembobotan *TF-IDF*, dan perhitungan similarity menggunakan Cosine Similarity.	Sistem mampu menghitung tingkat kesamaan judul dengan cepat. Contoh perhitungan menunjukkan *similarity* 20,84% antara dua judul contoh. Sistem juga menampilkan hasil dalam persentase dan daftar judul yang mirip.
5	(Halim and Lasut, 2024)	Aplikasi Deteksi Plagiarisme Dokumen Menggunakan Metode *TF-IDF* dan *Cosine Similarity* Berbasis *Web*	*TF-IDF, Cosine Similarity, Text Preprocessing (Lowercasing, Noise Removal, Stopword Removal, Tokenization, Feature Selection), Waterfall SDLC*	Proses dimulai dari pembersihan dan normalisasi teks, dilanjutkan dengan perhitungan bobot kata, kemudian diukur tingkat kemiripannya dengan dokumen dalam basis data. Sistem memberikan persentase	Hasil perhitungan manual sebesar 4% dan hasil aplikasi 4,34% menunjukkan konsistensi perhitungan yang baik.

				kemiripan sebagai output untuk memudahkan evaluasi keaslian dokumen.	

Analisis Tabel 2.5 menunjukkan bahwa meskipun TF-IDF dan *Cosine Similarity* telah banyak diaplikasikan, mayoritas penelitian terdahulu memiliki kelemahan dalam aspek validasi kebenaran data (*ground truth*). Evaluasi pada penelitian sebelumnya umumnya hanya membandingkan hasil perhitungan sistem dengan interpretasi subjektif peneliti secara manual, tanpa melibatkan otoritas akademik yang berwenang. Hal ini menyebabkan standar akurasi yang dihasilkan kurang merepresentasikan kebijakan institusi yang sebenarnya.

Oleh karena itu, penelitian ini mengisi celah tersebut dengan mengintegrasikan sistem secara langsung ke repositori institusi melalui *web scraping* otomatis, serta menerapkan metode evaluasi yang lebih kredibel. Penelitian ini melibatkan Koordinator Tugas Akhir sebagai validator ahli untuk menentukan *ground truth* terhadap kemiripan judul. Data dari pakar inilah yang kemudian digunakan untuk mengukur kinerja sistem melalui pengujian *Precision*, *Recall*, dan *F1-Score*. Dengan demikian, sistem yang dihasilkan tidak hanya akurat secara matematis, tetapi juga memiliki validitas yang sesuai dengan standar penilaian akademik di STMIK Palangkaraya.

BAB III

METODOLOGI PENELITIAN

3.1. Tinjauan Umum

Penelitian ini mengembangkan sistem pendeteksi kemiripan judul tugas akhir berbasis *web* yang terintegrasi langsung dengan repositori digital STMIK Palangkaraya untuk menciptakan solusi otomatis dalam pengecekan potensi kemiripan judul yang diajukan mahasiswa terhadap seluruh koleksi judul yang telah tersimpan, mengatasi keterbatasan proses manual yang saat ini masih diandalkan oleh dosen pembimbing dan koordinator tugas akhir. Sumber data penelitian diperoleh melalui mekanisme *web scraping* dari *website* perpustakaan institusi yang memuat 2174 judul tugas akhir hingga tahun 2024, menjamin kemutakhiran dataset tanpa ketergantungan pada input manual, dimana data judul kemudian diproses melalui tahapan preprocessing mencakup pembersihan dan normalisasi teks sebelum dilakukan pembobotan menggunakan metode *TF-IDF* dan perhitungan kemiripan dengan algoritma *Cosine Similarity* dengan pembagian tiga kategori yaitu Rendah (*similarity* di bawah 30%), Sedang (*similarity* 31-70%), dan Tinggi (*similarity* di atas 71%). Validasi kinerja sistem dilakukan melalui pendekatan ganda yang mengatasi keterbatasan penelitian terdahulu dengan melibatkan dosen secara langsung, dimana evaluasi akurasi algoritma mengukur metrik *Precision*, *Recall*, dan *F1-Score* terhadap ground truth yang ditetapkan dosen ahli, sedangkan evaluasi sistem sebagai produk dilakukan melalui uji *black-box* dengan dosen sebagai pengguna, sehingga sistem yang dibangun tidak hanya menjawab

kebutuhan akurasi teknis namun juga kesiapan operasional sebagai alat bantu keputusan yang telah divalidasi ahli dalam proses validasi judul tugas akhir di lingkungan STMIK Palangkaraya.

3.2. Jenis penelitian

Jenis penelitian dalam tugas akhir ini menggunakan pendekatan pengembangan perangkat lunak dengan menerapkan metode *Software Development Life Cycle* (SDLC) model *Waterfall*. Model *Waterfall* dipilih karena memiliki tahapan yang terstruktur dan berurutan, sehingga sesuai dengan kebutuhan pengembangan sistem yang telah ditentukan sejak awal. Tahapan tersebut meliputi: analisis kebutuhan, perancangan sistem, implementasi, pengujian, dan pemeliharaan. Melalui pendekatan ini, proses pengembangan dilakukan secara sistematis dan terdokumentasi, sehingga menghasilkan perangkat lunak yang stabil dan sesuai tujuan penelitian.

3.3. Desain Penelitian

Metode pengembangan yang digunakan dalam pembangunan sistem pendeteksi kemiripan judul tugas akhir ini adalah *Software Development Life Cycle* (SDLC) dengan model *Waterfall*. Tahapan dalam metode ini meliputi beberapa langkah berurutan sebagai berikut:

a. Analisis Kebutuhan

   Pada tahap ini dilakukan proses pengumpulan informasi mengenai kebutuhan sistem. Analisis kebutuhan diperoleh melalui studi terhadap proses verifikasi judul yang berlangsung di STMIK Palangkaraya, studi dokumentasi terhadap

website perpustakaan kampus, serta membaca literatur terkait teknik pengolahan teks, *TF-IDF*, dan *Cosine Similarity*. Dari analisis tersebut ditentukan kebutuhan fungsional dan nonfungsional yang menjadi dasar pengembangan sistem.

b. Perancangan Sistem (*Design*)

Tahap ini mencakup pembuatan rancangan arsitektur sistem, alur proses kerja aplikasi, serta desain antarmuka pengguna. UML digunakan sebagai alat bantu pemodelan, seperti *use case* diagram dan activity diagram. Selain itu, rancangan tampilan *website* juga disusun untuk menggambarkan struktur dan fitur utama sistem pendeteksi kemiripan judul.

c. Implementasi (*Coding*)

Tahap implementasi merupakan proses mengubah seluruh desain sistem menjadi kode program. Mengacu pada model *Waterfall*, kode program yang dibuat pada tahap ini masih berupa modul-modul terpisah yang nantinya akan diintegrasikan menjadi satu sistem yang lengkap. Pada penelitian ini, implementasi dilakukan dengan membangun modul-modul berikut:

1. Modul preprocessing teks
2. Modul pembobotan *TF-IDF*
3. Modul perhitungan *Cosine Similarity*
4. Modul *web scraping* untuk memperbarui dataset judul dari website perpustakaan
5. Modul antarmuka website untuk input judul dan menampilkan hasil kemiripan

d. Pengujian (*Testing*)

Pada tahap pengujian dilakukan pemeriksaan terhadap fungsi-fungsi sistem untuk memastikan bahwa setiap modul yang telah diintegrasikan bekerja sesuai dengan rancangan.

e. Verifikasi

Validasi sistem dilakukan melalui uji black-box yang melibatkan dosen. Validasi akurasi dilaksanakan dengan membandingkan output sistem terhadap ground truth dari penilaian tiga dosen ahli, menghasilkan metrik *Precision*, *Recall*, dan *F1-Score*. Validasi utilitas dilakukan melalui observasi penggunaan sistem oleh dosen dalam simulasi kerja nyata, menghasilkan data efisiensi waktu dan skor usabilitas.

f. Pemeliharaan (*Maintenance*)

Tahap pemeliharaan dilakukan setelah sistem selesai dibangun dan dinyatakan berfungsi dengan baik. Pada tahap ini dilakukan perbaikan *bug* yang mungkin masih muncul setelah sistem digunakan. Selain itu, pemeliharaan juga mencakup proses *web scraping* secara bertahap untuk penambahan data baru pada *website* perpustakaan, sehingga sistem tetap memiliki informasi yang terbaru dan relevan.

3.4. Instrumen Penelitian

Instrumen penelitian yang digunakan dalam tugas akhir ini merupakan cara yang digunakan untuk mengumpulkan data yang diperlukan dalam penelitian.

a. Studi Dokumentasi

Studi dokumentasi dilakukan dengan mempelajari data judul tugas akhir yang tersedia pada *website* perpustakaan kampus. Informasi seperti judul, penulis, tahun, dan akses link digunakan sebagai dataset utama dalam proses pengembangan model pendeteksi kemiripan.

b. *Web Scraping*

Data judul tugas akhir diperoleh melalui teknik *web scraping* pada website perpustakaan kampus. Data yang dikumpulkan meliputi judul, penulis, tahun, dan tautan akses, yang selanjutnya digunakan sebagai dataset utama dalam pengembangan model pendeteksi kemiripan. Teknik *web scraping* dipilih karena memungkinkan pengumpulan data secara otomatis dan efisien dari sumber yang tersedia secara daring.

3.5. Analisis Kebutuhan

Analisis kebutuhan dalam penelitian ini, yaitu perangkat keras dan perangkat lunak yang dibutuhkan untuk mendukung kelancaran pelaksanaan dan pengolahan data.

a. Perangkat Keras (*Hardware*)

Perangkat keras merupakan komponen fisik yang digunakan untuk mendukung proses pengembangan dan pengujian sistem dalam penelitian ini. Spesifikasi perangkat keras yang digunakan disesuaikan dengan kebutuhan pemrosesan data dan pengoperasian sistem pendeteksi kemiripan judul. Rincian perangkat keras yang digunakan ditunjukkan pada Tabel 3.1.

Tabel 3.1 Perangkat Keras

No	Perangkat Keras	Spesifikasi
1.	Laptop	Lenovo Ideapad 5
2.	Processor	AMD Ryzen 4500
3.	RAM	8.00 GB
4.	Storage	512 Gb

b. Perangkat Lunak (*Software*)

Perangkat lunak merupakan komponen pendukung yang digunakan dalam proses pengembangan, pengujian, dan pengoperasian sistem pendeteksi kemiripan judul tugas akhir. Perangkat lunak yang digunakan meliputi sistem operasi, bahasa pemrograman, serta *tools* pendukung lainnya. Rincian perangkat lunak yang digunakan ditunjukkan pada Tabel 3.2.

Tabel 3.2 Perangkat Lunak

No	Perangkat Lunak	Keterangan
1.	Windows 11	Sebagai Sistem Operasi
2.	Javascript	Sebagai bahasa pemrograman
3.	Google Spreadsheets	Sebagai *Database*
4.	Google Drive	Penyimpanan *Cloud*
4.	Python 3	Sebagai *web scraping*
5.	Figma	Sebagai perancangan (*interface*)

3.5.1. Analisis Data

Dalam melakukan analisis data diperlukan data yang valid dan dapat dipertanggungjawabkan. Oleh karena itu, penulis melakukan analisis berdasarkan hasil perhitungan menggunakan rumus yang digunakan serta kajian referensi resmi. Analisis data pada penelitian ini dilakukan melalui perhitungan manual untuk memverifikasi hasil sistem. Kategori kemiripan judul ditunjukkan pada Tabel 3.3. merujuk pada penelitian yang dilakukan oleh (Lumbansiantar et al., 2023)

Tabel 3.3 Kategori Tingkat Kemiripan Judul

Rentang Nilai	Kategori	Interpretasi
0-30%	Rendah	Judul memiliki perbedaan signifikan dalam metode, objek, atau fokus penelitian. Judul dapat digunakan.
31-70%	Sedang	Judul memiliki kesamaan dalam beberapa aspek seperti metode atau objek penelitian. Perlu pengecekan lebih lanjut dan klarifikasi perbedaan.
71-100%	Tinggi	Judul memiliki kesamaan yang sangat signifikan. Perlu revisi judul untuk memastikan originalitas penelitian.

Contoh soal

Q : Penerapan Metode *Fuzzy Mamdani* Untuk Diagnosa Penyakit Pata Tanaman Peria

D1 : Sistem Pakar Diagnosa Penyakit Pada Tanaman Jeruk Siam Banjar Menggunakan Metode *Fuzzy Mamdani*

D2 : Sistem Pakar Diagnosa Hama Dan Penyakit Pada Tanaman Sukun Menggunakan Metode *Certainty Factor*

a. *Preprocessing*

1. *Lowercasing*

Tahap pertama pada proses *preprocessing* adalah mengubah seluruh judul tugas akhir menjadi huruf kecil (*lowercase*) untuk menyeragamkan bentuk data. Proses ini bertujuan agar sistem tidak membedakan antara huruf besar dan huruf kecil dalam tahap analisis selanjutnya. Hasil proses case folding pada judul tugas akhir ditunjukkan pada Tabel 3.4.

Tabel 3.4 Hasil *Case Folding* pada Judul Tugas Akhir

Teks Asli	Hasil
Penerapan Metode Fuzzy Mamdani Untuk Diagnosa Penyakit Pata Tanaman Peria	penerapan metode fuzzy mamdani untuk diagnosa penyakit pata tanaman peria
Sistem Pakar Diagnosa Penyakit Pada Tanaman Jeruk Siam Banjar Menggunakan Metode Fuzzy Mamdani	sistem pakar diagnosa penyakit pada tanaman jeruk siam banjar menggunakan metode fuzzy mamdani
Sistem Pakar Diagnosa Hama Dan Penyakit Pada Tanaman Sukun Menggunakan Metode Certainty Factor	sistem pakar diagnosa hama dan penyakit pada tanaman sukun menggunakan metode certainty factor

2. *Tokenizing*

Setelah dilakukan proses *lowercasing*, setiap judul tugas akhir kemudian dipecah menjadi *token* atau kata per kata. Proses ini bertujuan agar setiap kata dapat dianalisis secara individual pada tahap selanjutnya. Hasil proses *tokenizing* pada judul tugas akhir ditunjukkan pada Tabel 3.5.

Tabel 3.5 Hasil *Tokenizing* pada Judul Tugas Akhir

Token	Hasil
penerapan metode fuzzy mamdani untuk diagnosa penyakit pata tanaman peria	[penerapan, metode, fuzzy, mamdani, untuk, diagnosa, penyakit, pata, tanaman, peria]
sistem pakar diagnosa penyakit pada tanaman jeruk siam banjar menggunakan metode fuzzy mamdani	[sistem, pakar, diagnosa, penyakit, pada, tanaman, jeruk, siam, banjar, menggunakan, metode, fuzzy, mamdani]
sistem pakar diagnosa hama dan penyakit pada tanaman sukun menggunakan metode certainty factor	[sistem, pakar, diagnosa, hama, dan, penyakit, pada, tanaman, sukun, menggunakan, metode, certainty, factor]

3. *Stopwords Removal*

Pada tahap ini, token yang mengandung kata-kata umum (*stopwords*) dihapus agar hanya kata-kata yang relevan dengan konteks judul yang dipertahankan. Proses ini bertujuan untuk meningkatkan kualitas analisis kemiripan. Hasil proses *stopword removal* pada judul tugas akhir ditunjukkan pada Tabel 3.6.

Tabel 3.6 Hasil *Stopword Removal* pada Judul Tugas Akhir

Tokens	Hasil
[penerapan, metode, fuzzy, mamdani, untuk, diagnosa, penyakit, pata, tanaman, peria]	[penerapan, metode, fuzzy, mamdani, diagnosa, penyakit, pata, tanaman, peria]
[sistem, pakar, diagnosa, penyakit, pada, tanaman, jeruk, siam, banjar, menggunakan, metode, fuzzy, mamdani]	[sistem, pakar, diagnosa, penyakit, tanaman, jeruk, siam, banjar, metode, fuzzy, mamdani]
[sistem, pakar, diagnosa, hama, dan, penyakit, pada, tanaman, sukun, menggunakan, metode, certainty, factor]	[sistem, pakar, diagnosa, hama, penyakit, tanaman, sukun, metode, certainty, factor]

4. *Stemming*

Setelah proses *stopword removal* dilakukan, *token* yang tersisa kemudian diproses untuk memperoleh bentuk kata dasarnya melalui proses *stemming*. Tahap ini bertujuan menyederhanakan kata berimbuhan menjadi kata dasar agar perhitungan *TF-IDF* menjadi lebih akurat. Hasil proses *stemming* pada judul tugas akhir ditunjukkan pada Tabel 3.7.

Tabel 3.7 Hasil *Stemming* pada Judul Tugas Akhir

Tokens	Hasil
[penerapan, metode, fuzzy, mamdani, diagnosa, penyakit, pata, tanaman, peria]	[terap, metode, fuzzy, mamdani, diagnosa, sakit, pata, tanam, peria]
[sistem, pakar, diagnosa, penyakit, tanaman, jeruk, siam, banjar, metode, fuzzy, mamdani]	[sistem, pakar, diagnosa, sakit, tanam, jeruk, siam, banjar, metode, fuzzy, mamdani]
[sistem, pakar, diagnosa, hama, penyakit, tanaman, sukun, metode, certainty, factor]	[sistem, pakar, diagnosa, hama, sakit, tanam, sukun, metode, certainty, factor]

b. *TF*

Tahap *Term Frequency* (TF) digunakan untuk menghitung frekuensi kemunculan setiap *term* pada masing-masing dokumen judul tugas akhir. Nilai *TF* menunjukkan seberapa sering suatu kata muncul dalam satu dokumen dan menjadi dasar dalam perhitungan bobot *TF-IDF*. Frekuensi kemunculan *term* pada setiap dokumen dengan rumus persamaan (2.1) ditunjukkan pada Tabel 3.8.

Tabel 3.8 Frekuensi Kemunculan *Term* (TF) pada Setiap Dokumen

Tokens	TF (Q)	TF (D1)	TF (D2)
terap	1	0	0
metode	1	1	1
fuzzy	1	1	0
mamdani	1	1	0
diagnosa	1	1	1
sakit	1	1	1
pata	1	0	0
tanam	1	1	1
peria	1	0	0
sistem	0	1	1
pakar	0	1	1
jeruk	0	1	0
siam	0	1	0
banjar	0	1	0

Tokens	TF (Q)	TF (D1)	TF (D2)
hama	0	0	1
sukun	0	0	1
certainty	0	0	1
factor	0	0	1

c. *IDF*

Tahap *Inverse Document Frequency* (IDF) digunakan untuk mengukur tingkat kepentingan suatu term dalam keseluruhan dokumen. Nilai IDF akan semakin besar jika suatu term jarang muncul pada dokumen lain, sehingga memberikan bobot lebih pada kata yang bersifat spesifik. Hasil perhitungan nilai IDF untuk setiap term dihitung dengan persamaan (2.2) ditunjukkan pada Tabel 3.9.

Tabel 3.9 Perhitungan Nilai IDF untuk Setiap Term

Tokens	*DF*	*IDF*
terap	1	1.477
metode	3	1
fuzzy	2	1.176
mamdani	2	1.176
diagnosa	3	1
sakit	3	1
pata	1	1.477
tanam	3	1
peria	1	1.477
sistem	2	1.176
pakar	2	1.176
jeruk	1	1.477
siam	1	1.477
banjar	1	1.477
hama	1	1.477
sukun	1	1.477
certainty	1	1.477
factor	1	1.477

d. TF-IDF

Tahap TF-IDF merupakan proses pembobotan term dengan mengombinasikan nilai *Term Frequency* (TF) dan *Inverse Document Frequency* (IDF). Pembobotan ini bertujuan untuk menonjolkan kata-kata yang penting dalam suatu dokumen serta mengurangi pengaruh kata yang sering muncul pada banyak dokumen. Nilai bobot TF-IDF setiap term pada masing-masing dokumen ditunjukkan pada Tabel 3.10.

Tabel 3.10 Nilai Bobot TF-IDF Setiap *Term* pada Masing-Masing Dokumen

Tokens	Q	D1	D2
terap	1.477	0	0
metode	1	1	1
fuzzy	1.176	1.176	0
mamdani	1.176	1.176	0
diagnosa	1	1	1
sakit	1	1	1
pata	1.477	0	0
tanam	1	1	1
peria	1.477	0	0
sistem	0	1.176	1.176
pakar	0	1.176	1.176
jeruk	0	1.477	0
siam	0	1.477	0
banjar	0	1.477	0
hama	0	0	1.477
sukun	0	0	1.477
certainty	0	0	1.477
factor	0	0	1.477

e. *Cosine Similarity*

Proses perhitungan *Cosine Similarity* dapat dihitung dengan menghitung *dot product* terlebih dahulu dengan persamaan (2.5).

Dot Product $A \cdot B$

metode: 1 × 1 = 1

fuzzy: 1.176 × 1.176 = 1.383

mamdani: 1.176 × 1.176 = 1.383

diagnosa: 1 × 1 = 1

sakit: 1 × 1 = 1

tanam: 1 × 1 = 1

Total Dot Product = 6.766

Persamaan (2.6) digunakan untuk menghitung panjang vektor.

Magnitude ||Q|| =

$$\sqrt{1{,}477^2 + 1^2 + 1{,}176^2 + 1{,}176^2 + 1^2 + 1^2 + 1{,}477^2 + 1^2 + 1{,}477^2} = 3.648$$

Magnitude ||D1|| =

$$\sqrt{1^2 + 1{,}176^2 + 1{,}176^2 + 1^2 + 1^2 + 1^2 + 1{,}176^2 + 1{,}176^2 + 1{,}477^2 + 1{,}477^2 + 1{,}477^2}$$

=4,009

Cosine Similarity (Q,D1) = $\frac{6{,}766}{(3{,}648 \times 4{,}009)} = 0{,}4626 = 46{,}3\%$

Dot Product $A \cdot B$ antara Q dan D2

metode: 1 × 1 = 1

diagnosa: 1 × 1 = 1

sakit: 1 × 1 = 1

tanam: 1 × 1 = 1

Total Dot Product = 4

Magnitude ||D2|| =

$$\sqrt{1^2 + 1^2 + 1^2 + 1^2 + 1{,}176^2 + 1{,}176^2 + 1{,}477^2 + 1{,}477^2 + 1{,}477^2 + 1{,}477^2}$$

=3,936

Persamaan (2.4) kembali digunakan untuk menghitung *Cosine Similarity*.

Cosine Similarity (Q,D2) = $\frac{4}{(3{,}648 \times 3{,}936)} = 0{,}2785 = 27{,}9\%$

Tahap *Cosine Similarity* digunakan untuk mengukur tingkat kemiripan antar judul tugas akhir berdasarkan vektor bobot TF-IDF. Hasil nilai *Cosine Similarity* berada pada rentang 0 hingga 1, di mana nilai yang semakin mendekati 1 menunjukkan tingkat kemiripan yang semakin tinggi. Hasil akhir perhitungan *Cosine Similarity* ditunjukkan pada Tabel 3.11.

Tabel 3.11 Hasil Akhir *Cosine Similarity*

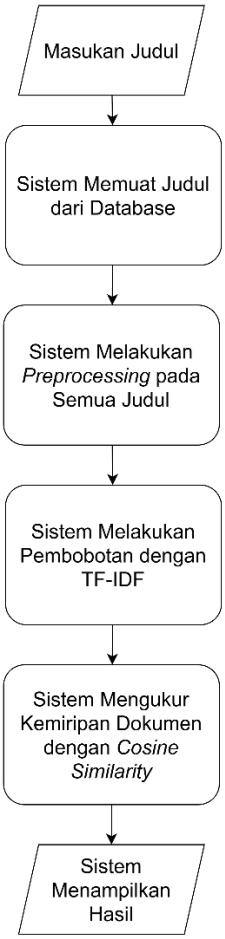
Dokumen	*Cosine Similarity*	Persentase Kemiripan	Kategori
D1	0.463	46.3%	Sedang
D2	0.279	27.9%	Rendah

3.5.2. Analisis Proses

Analisis proses pada sistem ini dibagi menjadi dua bagian, yaitu alur kerja website dan alur kerja *web scraping*. Alur kerja website menjelaskan bagaimana

sistem memproses input judul hingga menghasilkan nilai kemiripan, sedangkan alur kerja *web scraping* menggambarkan proses pengambilan dan pembaruan data judul dari *website* perpustakaan. Kedua proses ini saling mendukung dalam menghasilkan sistem yang berfungsi secara optimal.

a. Proses Alur Kerja Website

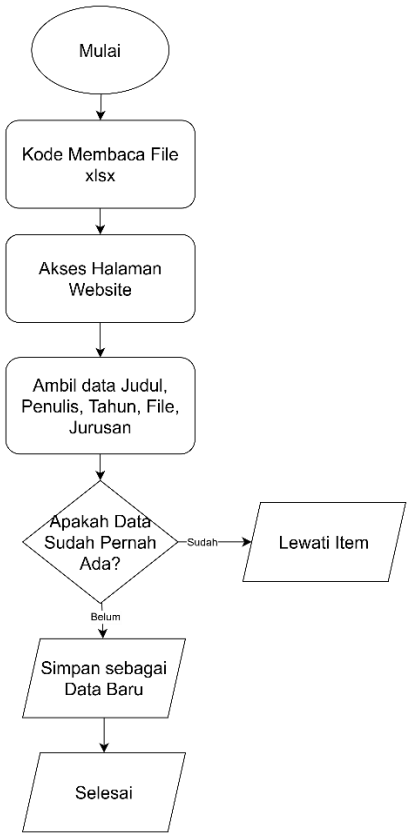


Gambar 3.1 *Flowchart* Alur Kerja *Website*

Proses alur kerja sistem pendeteksi kemiripan judul dimulai ketika pengguna memasukkan judul Tugas Akhir ke dalam *form* pada *website*. Setelah menerima input tersebut, sistem memuat seluruh judul yang tersimpan pada

database digital untuk dijadikan pembanding. Setiap judul, termasuk judul input pengguna, kemudian diproses melalui tahap preprocessing yang meliputi *case folding, tokenizing, stopword removal*, dan *stemming* agar seluruh teks berada dalam bentuk yang seragam. Selanjutnya, sistem menghitung bobot setiap kata menggunakan metode TF-IDF untuk mengubah setiap judul menjadi vektor numerik. Berdasarkan vektor tersebut, tingkat kemiripan antara judul pengguna dan seluruh judul pada database dihitung menggunakan algoritma *Cosine Similarity*. Pada tahap akhir, sistem menampilkan hasil berupa persentase kemiripan dan daftar judul yang paling relevan, sehingga pengguna dapat mengetahui apakah judul yang diajukan memiliki tingkat kesamaan dengan penelitian sebelumnya.

b. Proses Alur Kerja *Web Scraping*



Gambar 3.2 *Flowchart* Alur Kerja *Scraping*

Proses *web scraping* dilakukan untuk mengambil data judul Tugas Akhir secara otomatis dari website Perpustakaan STMIK Palangkaraya. Pada tahap ini, sistem terlebih dahulu memuat data lama dari file Excel untuk mengetahui judul mana yang sudah pernah tersimpan, sehingga proses scraping berikutnya tidak menghasilkan duplikasi. Selanjutnya, sistem mengakses setiap halaman kategori Tugas Akhir yang telah ditentukan dan mengekstraksi informasi penting seperti judul, penulis, tahun, NIM, jurusan, serta tautan file menggunakan teknik *parsing* HTML melalui *library BeautifulSoup*. Jika halaman memiliki *pagination*, sistem akan mengikuti tautan halaman

berikutnya hingga seluruh data berhasil diambil atau sampai batas duplikasi terpenuhi. Setiap data baru yang ditemukan kemudian disimpan dan ditambahkan ke dalam file Excel sebagai database utama yang digunakan oleh website pendeteksi kemiripan judul. Dengan cara ini, proses pembaruan data dapat dilakukan secara konsisten tanpa harus memasukkan data secara manual.

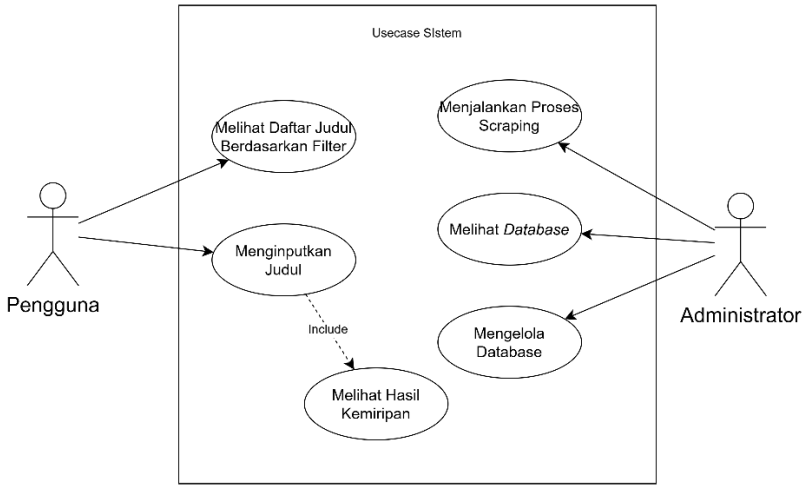
3.5.3. Analisis Kelemahan

Salah satu kelemahan utama dalam sistem ini adalah proses pengambilan data melalui *web scraping* masih perlu dijalankan setiap kali data ingin diperbarui, meskipun penyimpanan hasil *scraping* berjalan otomatis.

3.6. Desain

3.6.1. Desain Proses

a. *Usecase Diagram*

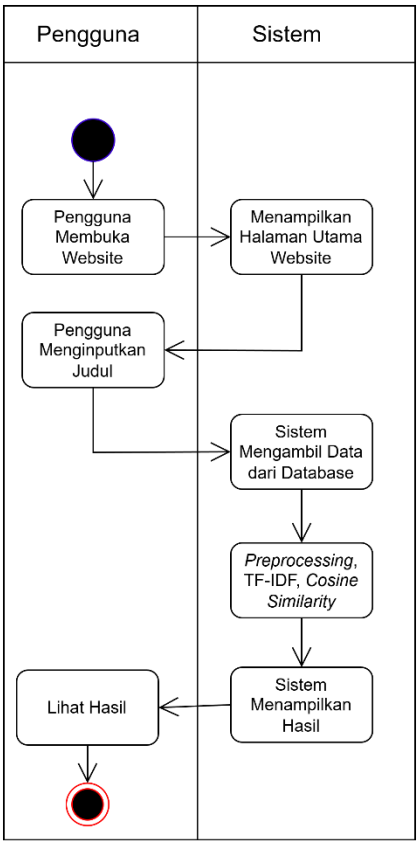


Gambar 3.3 *Usecase Diagram* proses

Pada sistem pendeteksi kemiripan judul ini terdapat dua aktor, yaitu Pengguna dan Administrator. Pengguna dapat melihat daftar judul berdasarkan

filter, serta menginputkan judul yang ingin diperiksa. Setelah judul dimasukkan, sistem secara otomatis menampilkan hasil kemiripan sebagai bagian dari proses tersebut. Sementara itu, Administrator bertanggung jawab dalam menjaga ketersediaan dan pembaruan data melalui proses *web scraping*, pengelolaan isi basis data, serta pemantauan data yang telah tersimpan.

b. *Activity Diagram*



Gambar 3.4 Activity Diagram Alur Sistem

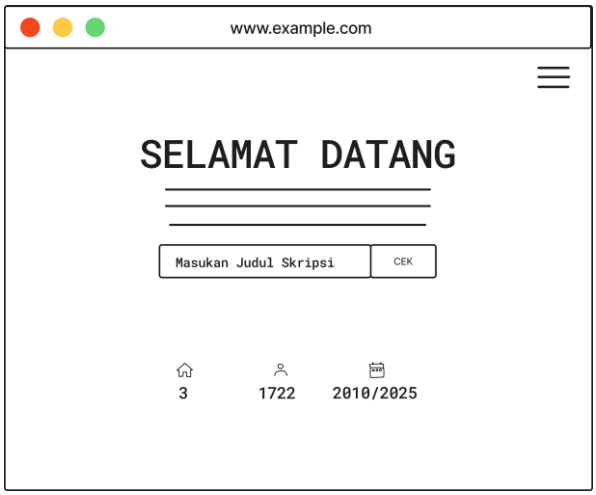
Activity diagram menggambarkan alur proses ketika pengguna melakukan pengecekan kemiripan judul melalui *website*. Proses dimulai saat pengguna membuka halaman utama *website*, kemudian sistem menampilkan

antarmuka awal yang berisi *form input* judul. Setelah halaman tersebut ditampilkan, pengguna memasukkan judul Tugas Akhir yang ingin diperiksa.

Setelah judul diterima, sistem mengambil seluruh data judul pembanding dari database untuk digunakan dalam proses perhitungan. Selanjutnya, sistem melakukan serangkaian proses pemrosesan teks, seperti *Case Folding*, *Tokenizing*, *Stopwords Removal* dan *Stemming*, perhitungan TF-IDF, dan *Cosine Similarity* untuk menentukan tingkat kemiripan judul yang diinputkan pengguna dengan judul-judul yang tersedia dalam database. Setelah perhitungan selesai, sistem menyajikan hasil berupa persentase kemiripan dan daftar judul terkait kepada pengguna. Proses berakhir ketika pengguna melihat hasil yang ditampilkan oleh sistem.

3.6.2. Desain Perangkat

a. Halaman Utama

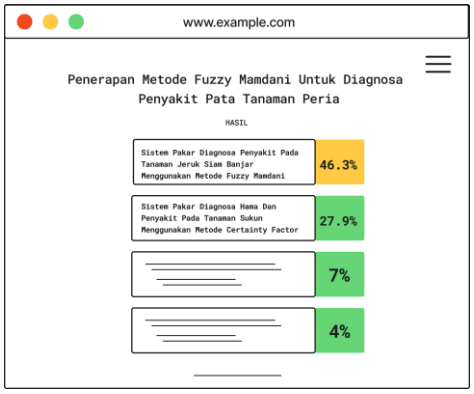


Gambar 3.5 Desain Halaman Utama

Halaman utama dirancang sebagai halaman pertama yang dilihat oleh pengguna saat mengakses sistem. Pada bagian atas halaman ditampilkan tulisan “Selamat Datang” sebagai sapaan awal, diikuti oleh sub judul yang menjelaskan tujuan sistem, yaitu membantu pengguna mengecek kemiripan judul tugas akhir. Di bagian tengah halaman tersedia kolom input judul tempat pengguna memasukkan judul yang ingin diperiksa, disertai dengan tombol “Cari” untuk memulai proses pengecekan.

Selain itu, halaman utama juga menampilkan sejumlah statistik ringkas yang memberikan gambaran mengenai isi database sistem, seperti jumlah total program studi yang terdata, jumlah keseluruhan judul tugas akhir yang tersedia, serta rentang tahun dari data yang terkumpul. Informasi ini ditampilkan dalam bentuk kartu ringkas untuk memberikan konteks kepada pengguna mengenai cakupan data yang digunakan dalam proses pendeteksian kemiripan.

b. Halaman Hasil

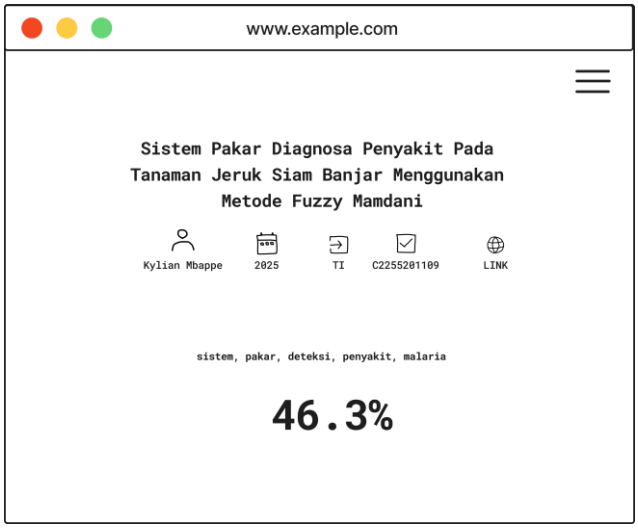


Gambar 3.6 Desain Halaman Hasil

Halaman hasil merupakan bagian yang menampilkan keluaran dari proses pengecekan kemiripan judul. Setelah pengguna memasukkan judul dan menekan tombol “Cari”, sistem akan menampilkan daftar judul yang paling mirip dengan judul yang dimasukkan. Daftar ini disajikan dalam bentuk list berisi maksimal 10 judul yang memiliki tingkat kemiripan tertinggi berdasarkan perhitungan TF-IDF dan *Cosine Similarity*.

Setiap judul dalam daftar ditampilkan bersama persentase kemiripan, sehingga pengguna dapat langsung mengidentifikasi apakah judul mereka tergolong aman atau memiliki kemiripan yang tinggi dan perlu direvisi.

c. Halaman Detail



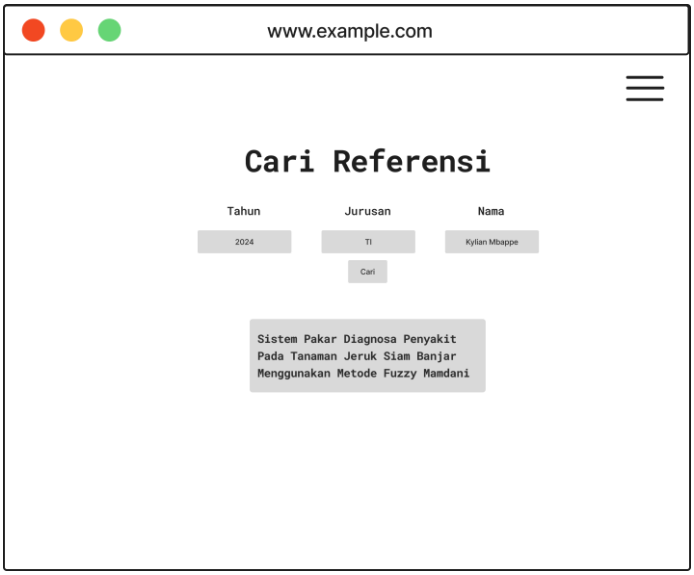
Gambar 3.7 Desain Halaman Detail

Halaman detail dirancang untuk menampilkan informasi lengkap dari salah satu judul yang dianggap mirip berdasarkan hasil perhitungan sistem. Pada halaman ini, pengguna dapat melihat judul tugas akhir yang

serupa beserta informasi penting lainnya seperti nama penulis, tahun publikasi, program studi, dan NIM. Selain itu, halaman ini juga menyediakan tautan langsung ke halaman perpustakaan kampus yang diperoleh dari proses *web scraping*, sehingga pengguna dapat melihat sumber aslinya secara lebih lengkap.

Sistem juga menampilkan persentase kemiripan antara judul yang diperiksa dengan judul yang ditampilkan, serta daftar kata-kata yang sama atau memiliki kesesuaian tinggi yang berkontribusi pada nilai *similarity* tersebut. Dengan tampilan detail ini, pengguna dapat memahami secara jelas faktor apa yang membuat kedua judul serupa dan menentukan apakah judul yang mereka ajukan perlu diperbaiki atau diganti.

d. Halaman Daftar Judul



Gambar 3.8 Desain Halaman Daftar Judul

Halaman daftar judul berfungsi sebagai pusat pencarian referensi bagi pengguna. Pada halaman ini ditampilkan seluruh judul tugas akhir yang ada didatabase, yang dapat diakses dan disaring menggunakan beberapa filter pencarian. Pengguna dapat melakukan penyaringan berdasarkan tahun, jurusan, maupun nama penulis untuk memudahkan proses menemukan referensi yang relevan.

Setelah filter diterapkan, sistem akan menampilkan daftar judul sesuai kriteria yang dipilih dalam bentuk tabel atau list yang rapi. Setiap item pada daftar juga dapat diklik untuk membuka halaman detail sehingga pengguna bisa melihat informasi lengkap terkait judul tersebut. Dengan adanya halaman ini, pengguna dapat menelusuri kumpulan judul secara fleksibel dan efisien tanpa harus melakukan pencarian manual satu per satu.

3.6.3. Desain Basisdata

Basis data pada sistem ini menggunakan Google Spreadsheets sebagai media penyimpanan utama yang menampung seluruh data judul tugas akhir hasil proses *web scraping* dari *website* Perpustakaan STMIK Palangkaraya. Struktur tabel dirancang secara sederhana namun telah mencakup seluruh informasi yang dibutuhkan dalam proses pendeteksian kemiripan judul. Setiap baris pada tabel merepresentasikan satu entri judul tugas akhir, sedangkan setiap kolom berisi atribut yang relevan. Desain kolom pada basis data ditunjukkan pada Tabel 3.12.

Tabel 3.12 Database Spreadsheets

Nama Kolom	Tipe Data	Keterangan
ID	INT	Berisikan ID judul
Judul	TEXT	Judul Tugas Akhir
NIM	VARCHAR	Nama Penulis/Pengarang
Tahun	INT	NIM Mahasiswa
Penulis	VARCHAR	Tahun Penyusunan
Jurusan	VARCHAR	TI/SI/MI
File	LINK	Link Halaman Perpustakaan Hasil Scraping

3.7. Jadwal Penelitian

Tabel 3.13 Jadwal Penelitian

No.	Tahapan	Oktober 2025				November 2025				Desember 2025				Januari 2026				Febuari 2026				Maret 2026			
		I	II	III	IV	I	II	III	IV	I	II	III	IV	I	II	III	IV	I	II	III	IV	I	II	III	IV
1	Identifikasi Masalah																								
3	Analisis Kebutuhan sistem																								
4	Pengumpulan data																								
5	Analisis data																								
6	Pembuatan Desain																								
7	Impelementasi																								
8	Penyusuan Laporan Tugas Akhir																								

DAFTAR PUSTAKA

Amanah Amnun Zulfa, Tatang Ibrahim, O.A., 2025. Peran Sistem Informasi Akademik Berbasis Web Dalam Upaya Meningkatkan Efektivitas Dan Efisiensi Pengelolaan Akademik Di Perguruan Tinggi. Jurnal Tahsinia 6, 115–134.

Dikilitaş, Y., Çakal, Ç., Okumuş, A.C., Yalçın, H.N., Yıldırım, E., Ulusoy, Ö.F., Macit, B., Kırkaya, A.E., Yalçın, Ö., Erdoğmuş, E., Sayar, A., 2024. Performance Analysis for Web Scraping Tools: Case Studies on Beautifulsoup, Scrapy, Htmlunit and Jsoup. Lecture Notes in Networks and Systems 960, 471–480. https://doi.org/10.1007/978-3-031-56728-5_39

Eko, H., Adi, U., Nasiri, A., Informatika, M.T., Framework, M., 2023. Studi Dan Analisis Algoritmakemiripan Cosine Similarity Untuk Mengidentifikasikesamaan Judul Skripsi. PROSIDING SEMINAR ILMIAH SISTEM INFORMASI DAN TEKNOLOGI INFORMASI XII, 1–9.

Fandopa, J.A., Santoso, N., 2022. Pengembangan Sistem Informasi Manajemen Percetakan pada Gajayana Digital Printing Kota Malang berbasis Website. Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer 6, 5371–5379.

Halim, J., Lasut, D., 2024. Document Plagiarism Detection Application Using Web-Based TF-IDF and Cosine Similarity Methods. bit-Tech 7, 202–213. https://doi.org/10.32877/bt.v7i2.1697

Hendrawan, E., Meisel, M., Sari, D.N., 2023. Analysis And Implementation Of Computer Network Systems Using Software Draw.io. Asia Information System Journal 2, 9–15.

Kurniawan, S., Wiranata, W., Kusnan, K., Ma'muriyah, N., Ting, V.V., 2023. Pemanfaatan Komputasi Awan (Cloud Computing) Pada Bidang Pendidikan. Journal of Information System and Technology 4, 403–405. https://doi.org/10.37253/joint.v4i2.6243

Lumbansiantar, S., Dwiasnati, S., Fatonah, N.S., 2023. Penerapan Metode Cosine Similarity Dalam Mendeteksi Plagiarisme Pada Jurnal. Format : Jurnal Ilmiah Teknik Informatika 12, 142. https://doi.org/10.22441/format.2023.v12.i2.007

Meidelfi, D., Rahmayuni, I., Hidayat, T., Chandra, D., 2021. TF-IDF Implementation for Similarity Checker on The Final Project Title 3, 40–52.

Muhyidin, M.A., Sulhan, M.A., Sevtiana, A., 2020. Perancangan Ui/Ux Aplikasi My Cic Layanan Informasi Akademik Mahasiswa Menggunakan Aplikasi Figma. Jurnal Digit 10, 208. https://doi.org/10.51920/jd.v10i2.171

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 Fakultas Teknik Universitas Jabal Ghafur. Sagita Academia Journal 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

Salman, 2025. Sistem Pendeteksi Kemiripan Judul Skripsi Berbasis Web Menggunakan Nlp Dan Word Embeddings. Jurnal Pendidikan Indonesia:Teori, Penelitian dan Inovasi 5. https://doi.org/10.59818/jpi.v5i5.1964

Trilaksono, A.R., Husain, T., Doharma, R., 2020. Uji Model Keberhasilan Sistem Informasi: Media Penyimpanan Google Drive. Jurnal Teknologi Sistem Informasi dan Aplikasi 3, 57. https://doi.org/10.32493/jtsi.v3i2.4434

Fitria Nur Hasanah M.Pd, R. S. (2020). *Rekayasa Perangkat Lunak* (Pertama ed.). Sidoarjo: Universitas Muhammadiyah Sidoarjo.

Google. (2024). *Understanding spreadsheets: an essential guide*. Retrieved November 21, 2025, from https://workspace.google.com/blog/productivity-collaboration/understanding-spreadsheets-essential-guide

Halim, J. (2024). *Aplikasi Deteksi Plagiarisme Dokumen Dengan Metode Tf-idf Dan Cosine Similarity Berbasis Web.* Tangerang: Universitas Buddhi Dharma.

L A M P I R A N

SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER
(STMIK) PALANGKARAYA
Jl. G. Obos No.114 Telp.0536-3224593, 3225515 Fax.0536-3225515 Palangka Raya
email : humas@stmikplk.ac.id - website : www.stmikplk.ac.id

SURAT TUGAS PEMBIMBING TUGAS AKHIR

NO. 285/STMIK-3.C.2/AU/XI/2025

Ketua Program Studi Teknik Informatika Sekolah Tinggi Manajemen Informatika dan Komputer (STMIK) Palangkaraya menugaskan nama-nama tersebut di bawah ini :

1. Nama : Dr. Maura. Widyaningsih, S.Kom., M.Cs
   NIK : 197502232009002
   Sebagai Pembimbing I dalam **PEMBUATAN PROGRAM**
2. Nama : Ir. Abdul Hadi, S.T., M.Kom
   NIK : 198505102021101
   Sebagai Pembimbing II dalam **PENULISAN TUGAS AKHIR**

Untuk membimbing Tugas Akhir mahasiswa :

Nama	:	Ibnu Rofik
NIM	:	C2255201109
Program Studi	:	TEKNIK INFORMATIKA (55201)
Tanggal Daftar	:	05/11/2025
Judul Tugas Akhir	:	Penerapan TF-IDF dan Cosine Similarity pada Deteksi Kemiripan Judul Tugas Akhir di STMIK Palangkaraya

Demikian surat ini dibuat agar dapat dipergunakan sebagaimana mestinya, dan dilaksanakan dengan penuh tanggung jawab.

Palangka Raya, 07 November 2025

Ketua Program Studi,

November 2025

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

TEMBUSAN :
1. Pembimbing I dan II
2. Mahasiswa yang bersangkutan
3. Arsip Prodi TI

SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER
(STMIK) PALANGKARAYA
Jl. G. Obos No.114 Telp.0536-3224593, 3225515 Fax.0536-3225515 Palangkaraya
email : humas@stmikplk.ac.id - website : www.stmikplk.ac.id

KARTU KEGIATAN KONSULTASI TUGAS AKHIR

Nama Mahasiswa : Ibnu Rofik
NIM : C2255201109
Tanggal Persetujuan Judul : 8 November 2025
Judul Tugas Akhir : Penerapan TF-IDF dan cosine similarity pada deteksi kemiripan Judul tugas akhir di STMIK Palangka Raya

No.	Tanggal Konsultasi		Uraian	Tanda Tangan
	Terima	Kembali		
1	8-11-25		konsultasi judul.	
2	13-11-25		konsultasi Bab 1 & 2.	
3	17-11-25		konsultasi Bab 3	
4	12/12 25		Revisi Bab 3.	
5	17/12 25		Acc maju sempro	
6	13/11 25		konsultasi penulisan BAB I	
7	12/12 25		konsultasi penulisan BAB 2-3 Penomoran tabel, gambar	
8	16/12 25		konsultasi BAB III	
9	18/12 25		Narasi setiap gambar & tabel	
	19/12 25		Acc seminar	

Perihal : Permohonan Izin Penelitian dan Pengumpulan Data untuk Tugas Akhir

Kepada

Yth. Kepala UPT Perpustakaan STMIK Palangka Raya

Jl. G. Obos No.114, Palangka Raya, Kalimantan Tengah

Dengan Hormat,

Sehubungan dengan penyusunan Tugas Akhir mahasiswa sebagai persyaratan kelulusan Program Studi Teknik Informatika (S1) pada STMIK Palangkaraya, maka dengan ini kami sampaikan permohonan izin penelitian dan pengumpulan data bagi mahasiswa kami berikut:

Nama	: IBNU ROFIK
NIM	: C2255201109
Prodi (Jenjang)	: Teknik Informatika
Thn. Akad. (Semester)	: 2025/2026 (8)
Lama Penelitian	: 22 Januari 2026 s.d 22 Febuari 2026
Tempat Penelitian	: Website Perpustakaan STMIK Palangkaraya

Dengan judul Tugas Akhir:

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

Adapun ketentuan dan aturan pemberian informasi dan data yang diperlukan dalam penelitian tersebut menyesuaikan dengan ketentuan/peraturan pada instansi Bapak/Ibu

Demikian permohonan ini disampaikan, atas perhatian dan kerja samanya diucapkan terima kasih.

Palangka Raya, 22 Januari 2026

Peneliti,



IBNU ROFIK

NIM. C2255201109

SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER
(STMIK) PALANGKARAYA
Jl. G. Obos No.114 Telp.0536-3224593, 3225515 Fax.0536-3225515 Palangka Raya
email : humas@stmikplk.ac.id - website : www.stmikplk.ac.id

SURAT TUGAS
PENGUJI SEMINAR PROPOSAL TUGAS AKHIR

364/STMIK-3.C.2/AU/XII/2025

Ketua Program Studi Teknik Informatika Sekolah Tinggi Manajemen Informatika dan Komputer (STMIK) Palangkaraya menugaskan kepada nama-nama berikut :

1. Nama : Sulistyowati, S.Kom., M.Cs.
   NIK : 198212162007002
   Sebagai Ketua
2. Nama : Dr. Maura. Widyaningsih, S.Kom., M.Cs
   NIK : 197502232009002
   Sebagai Sekretaris
3. Nama : Ir. Abdul Hadi, S.T., M.Kom
   NIK : 198505102021101
   Sebagai Anggota

Untuk menguji seminar proposal tugas akhir mahasiswa :

Nama Mahasiswa	:	Ibnu Rofik
N I M	:	C2255201109
Hari / Tanggal Ujian	:	Selasa, 13 Januari 2026
Waktu	:	08.00 WIB
Ruangan	:	TI 4
Judul TA	:	Penerapan TF-IDF dan Cosine Similarity pada Deteksi Kemiripan Judul Tugas Akhir di STMIK Palangkaraya

Demikian surat ini dibuat agar dapat dipergunakan sebagaimana mestinya dan dilaksanakan dengan penuh tanggung jawab.

Palangka Raya, 27 Desember 2025

Ketua Program Studi Teknik Informatika,

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

Tembusan :
1. Dosen Penguji
2. Mahasiswa yang Bersangkutan
3. Arsip

SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER
(STMIK) PALANGKARAYA
Jl. G. Obos No.114 Telp.0536-3224593, 3225515 Fax.0536-3225515 Palangka Raya
email : humas@stmikplk.ac.id - website : www.stmikplk.ac.id

DAFTAR HADIR PESERTA
SEMINAR PROPOSAL TUGAS AKHIR

Nama Penyaji : Ibnu Rofik
Hari/ Tanggal : Selasa, 13 Januari 2026
Waktu : 08.00 WIB
Judul Proposal : Penerapan TF-IDF dan Cosine Similarity pada Deteksi Kemiripan Judul Tugas Akhir di STMIK Palangkaraya

No.	Nama Mahasiswa	NIM	Tanda Tangan
1	FERNANDO DAYADUTA	C2255201102	
2	I Wayan Nityananda Sudiartika	C2255201087	
3	Habib Abdel Wahab	C2255201121	
4	Agriva Abel Patewah	C2255201089	
5	Revinadiya Pitaloka	C2255201049	
6	Daniel Setiawan Purnomo	C2255201107	
7	Josepha Amanda V.P.	C2255201116	
8	I Kadek Deo W.	C2255201103	
9	Okdilia	C2255201112	
10	Megawati	C2255201170	
11	Ryan Achmad Antama	C2255201105	
12	NORHALISAH	C2255201024	
13	QODLI ZAKA	C2255201019	
14			
15			
16			
17			
18			
19			
20			
21			
22			
23			
24			
25			
26			
27			
28			
29			
30			

Palangka Raya, 13 Januari 2026

Mengetahui :
Ketua Tim Penguji,

Sulistyowati, S.Kom., M.Cs

Mahasiswa,

Ibnu Rofik

SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER
(STMIK) PALANGKARAYA
Jl. G. Obos No.114 Telp.0536-3224593, 3225515 Fax.0536-3225515 Palangka Raya
email : humas@stmikplk.ac.id - website : www.stmikplk.ac.id

BERITA ACARA
SEMINAR PROPOSAL TUGAS AKHIR

Periode : Januari 2026

Pada hari ini Selasa, 13 Januari 2026 Pukul 08.00 WIB sampai dengan 09.00 WIB telah dilaksanakan Seminar Proposal Tugas Akhir mahasiswa atas nama :

Nama Mahasiswa	:	Ibnu Rofik
N I M	:	C2255201109
Program Studi	:	Teknik Informatika
Tahun Angkatan	:	2022
Judul TA	:	Penerapan TF-IDF dan Cosine Similarity pada Deteksi Kemiripan Judul Tugas Akhir di STMIK Palangkaraya

Dosen Penguji :

	Nama		Nilai	Tanda Tangan
1.	Sulistyowati, S.Kom., M.Cs	=	.....	(.....)
2.	Dr. Maura. Widyaningsih, S.Kom., M.Cs	=	.....	(.....)
3.	Ir. Abdul Hadi, S.T., M.Kom	=	.....	(.....)

Hasil Ujian : **LULUS** / ~~**TIDAK LULUS**~~ *) NILAI = ...

Revisi (Tidak Ada / Minor / Mayor *)

Catatan Penting : 1. Lama Perbaikan : ..... hari (Maks. 15 hari)
2. Jika lebih dari 1 (satu) bulan dikenakan denda Rp. 300.000,- (Tiga Ratus ribu rupiah) per bulan dari tanggal ujian.
3. Jika lebih dari 3 (tiga) bulan dari tanggal ujian maka hasil ujian dibatalkan dan wajib mengajukan judul dan pembimbing baru. Wajib membayar Denda dan membayar biaya seminar ulang.

Palangka Raya, 13 Januari 2026

Ketua Penguji,	Mahasiswa,
Sulistyowati, S.Kom., M.Cs	**Ibnu Rofik**
NIK. 198212162007002	NIM. C2255201109

Mengetahui :
Ketua Program Studi Teknik Informatika,

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

Tembusan:
Mahasiswa yang bersangkutan.
Berita acara ini diberikan setelah menyelesaikan semua revisi dari semua penguji
**) Coret yang tidak perlu*