**KOMBINASI ALGORITMA *TF-IDF* DAN *FUZZY MATCHING* UNTUK DETEKSI KEMIRIPAN JUDUL SKRIPSI BERBASIS APLIKASI ANDROID.**

STUDI KASUS: JUDUL SKRIPSI PRODI TEKNIK INFORMATIKA IIB DARMAJAYA

**SKRIPSI**



**Disusun oleh :**

**Arif Nur Listanto 2011010035**

**PROGRAM STUDI TEKNIK INFORMATIKA**

**FAKULTAS ILMU KOMPUTER**

**INSTITUT INFORMATIKA DAN BISNIS DARMAJAYA**

**BANDAR LAMPUNG**

**TAHUN 2024**

# DAFTAR ISI

[DAFTAR ISI I](#_Toc170986522)

[DAFTAR GAMBAR III](#_Toc170986523)

[DAFTAR TABEL IV](#_Toc170986524)

[BAB I 1](#_Toc170986525)

[PENDAHULUAN 1](#_Toc170986526)

[1.1 Latar Belakang 1](#_Toc170986527)

[1.2 Indentifikasi Masalah 3](#_Toc170986528)

[1.3 Rumusan Masalah 3](#_Toc170986529)

[1.4 Batasan Masalah 3](#_Toc170986530)

[1.5 Tujuan Penelitian 4](#_Toc170986531)

[1.6 Manfaat Penelitian 4](#_Toc170986532)

[1.7 Sistematika Penulisan 5](#_Toc170986533)

[BAB II 6](#_Toc170986534)

[TINJAUAN PUSTAKA 6](#_Toc170986535)

[2.1 Implementasi 6](#_Toc170986536)

[2.2 Algoritma 6](#_Toc170986537)

[2.3 Deteksi 6](#_Toc170986538)

[2.4 Aplikasi Android 7](#_Toc170986539)

[2.5 Algoritma *TF-IDF* 8](#_Toc170986540)

[2.5.1 Term Frequency 8](#_Toc170986541)

[2.5.2 Inverse Document Frequency 9](#_Toc170986542)

[2.5.3 Langkah Langkah Algoritma *TF – IDF* 10](#_Toc170986543)

[2.6 Algoritma Fuzzy Matching 11](#_Toc170986544)

[2.6.1 Cara Kerja Algoritma Fuzzy Matching 11](#_Toc170986545)

[2.7 UML (Unified Modeling Language) 12](#_Toc170986546)

[2.7.1 Activity Diagram 12](#_Toc170986547)

[2.8 Database MySQL 13](#_Toc170986548)

[2.9 Pengujian *Black Box* 13](#_Toc170986549)

[2.10 Hasil Penelitian Sebelumnya 14](#_Toc170986550)

[BAB III METODOLOGI PENELITIAN 17](#_Toc170986551)

[3.1 Metode Pengumpulan Data 17](#_Toc170986552)

[3.1.1 Wawancara 17](#_Toc170986553)

[3.1.2 Pengamatan 17](#_Toc170986554)

[3.1.3 Tinjauan Pustaka 17](#_Toc170986555)

[3.2 Kombinasi Algoritma *TF-IDF* Dan *Fuzzy Matching* 17](#_Toc170986556)

[3.3 Analisis Sistem 18](#_Toc170986557)

[3.2.1 Analisis Kebutuhan Perangkat Lunak 18](#_Toc170986558)

[3.3 Perancangan Sistem 18](#_Toc170986559)

[3.4 Tahapan Membangun *Prototype* 20](#_Toc170986560)

[3.4.1 Penggambaran *Usecase Diagram* 20](#_Toc170986561)

[3.4.2 Penggambaran *Activity Diagram* 21](#_Toc170986562)

[3.5 Desain Aplikasi 22](#_Toc170986563)

[3.5.1 Desain Tampilan Awal Aplikasi 22](#_Toc170986564)

[3.6 Jadwal Penelitian 25](#_Toc170986565)

[BAB IV 26](#_Toc170986566)

[HASIL DAN PEMBAHASAN 26](#_Toc170986567)

[4.1 Hasil Penelitian 26](#_Toc170986568)

[4.1.1 Tampilan Halaman Utama 26](#_Toc170986569)

[4.1.2 Tampilan Halaman Cek Judul Skripsi 26](#_Toc170986570)

[4.1.3 Tampilan Halaman Tanya *Artificial Intelligence* 26](#_Toc170986571)

[DAFTAR PUSTAKA 27](#_Toc170986572)

# DAFTAR GAMBAR

[**Gambar 3.3.1** *Flowchart* Proses Pengeckan 19](#_Toc172090815)

[**Gambar 3.4.1** Diagram *Usecase* 20](#_Toc172090816)

[**Gambar 3.4.2** *Activity* Diagram Mahasiswa 21](#_Toc172090817)

[**Gambar 3.4.3** *Activity Diagram Admin* 22](#_Toc172090818)

[**Gambar 3.5.1** *Prototype* Tampilan Awal Aplikasi 23](#_Toc172090819)

[**Gambar 3.5.3** *Prototype* Tampilan Awal Admin 23](#_Toc172090820)

[**Gambar 3.5.4** *Prototype* Tampilan Tambah Judul 24](#_Toc172090821)

[**Gambar 3.5.5** *Prototype* Tampilan Edit Judul 24](#_Toc172090822)

# DAFTAR TABEL

[**Tabel 2.7.2** Simbol – Simbol *Activity Diagram* 12](#_Toc172091094)

[**Tabel 2.10.1** Tinjauan Pustaka Jurnal 14](#_Toc172091095)

[**Tabel 3.6.1** Jadwal Penelitian 25](#_Toc172091096)

# BAB I

# PENDAHULUAN

## 1.1 Latar Belakang

Skripsi merupakan salah satu tahapan akhir dalam menyelesaikan studi di perguruan tinggi. Setiap mahasiswa pada jalur skripsi harus menentukan tema dan judul skripsi sesuai dengan bidang ilmu masing-masing mahasiswa. Skripsi harus ditulis dengan orisinil dalam arti tidak plagiat dan tidak memiliki kermiripan yang sangat tinggi antara satu penelitian dengan penelitian lainnya.

Setiap judul skripsi yang diajukan oleh mahasiswa akan di filter oleh ketua Kelompok Bidang Keilmuan (KBK). Di Darmajaya khususnya prodi Teknik Informatika terdapat 3 konsenterasi bidang keilmuan yaitu Artificial Intelligence, Mobile Technology, dan Multimedia. Masing-masing bidang keilmua memiliki ketua KBK. KBK akan melakukan pengecekan dari setiap judul skripsi yang diajukan oleh mahasiswa dalam rangka menghindari judul identik, kemiripan yang sangat tinggi, dan plagiarisme dalam penelitian skripsi.

Masalah yang sering terjadi yaitu judul yang diajukan oleh mahasiswa seringkali identik, sangat mirip dengan judul yang sudah pernah diajukan sebelumnya, ataupun mahasiswa hanya mengganti judul dengan lokasi penelitian yang berbeda tanpa adanya nilai keterbaharuan dari penelitian tersebut. Judul skripsi yang identik, sangat mirip, dan plagiat pasti ditolak oleh KBK. Saat ini terdapat ±2148 judul skripsi pada *spreadsheet* prodi teknik informatika. Sehingga meyulitkan untuk mendeteksi kemiripan pada judul skripsi yang diajukan. Untuk mendeteksi kemiripan pada judul skripsi.

Kombinasi algoritma *TF-IDF* dengan *Fuzzy Matching* merupakan salah satu algoritma yang cocok untuk mendeteksi tingkat kemiripan judul skripsi yang diajukan oleh mahasiswa.

Algoritma *TF-IDF* adalah metode yang digunakan untuk mengukur seberapa penting sebuah kata dalam sebuah dokumen [1], dimana *TF-IDF* secara umum didefinisikan sebagai 2 algoritma yang berbeda. *TF (Term Frequence)* merupakan algoritma yang menghitung seberapa sering sebuah kata muncul dalam sebuah dokumen, sedangkan *IDF (Inverse Document Frequency)* merupakan algoritma yang menghitung seberapa jarang sebuah kata muncul dalam sebuah dokumen. Algoritma kombinasi lainnya yaitu *Fuzzy Matching* merupakan metode pencocokan string yang menggunakan pendekatan terhadap pola dari string yang dicari, dimana algoritma *Fuzzy Matching* bekerja dengan cara yang berbeda dari pencocokan string biasa (*exact matching*) [2], algoritma ini berfokus pada menemukan kecocokan string yang “mirip”. Sehingga apabila kedua algoritma ini dikombinasikan maka akan menghasilkan tingkat akurasi yang cukup tinggi dalam mendeteksi kemiripan judul – judul skripsi mahasiswa.

Berdasarkan uraian diatas, maka penulis mengajukan judul skripsi **“IMPLEMENTASI ALGORITMA *TF-IDF* DAN *FUZZY MATCHING* UNTUK DETEKSI KEMIRIPAN JUDUL SKRIPSI BERBASIS APLIKASI ANDROID. STUDI KASUS: JUDUL SKRIPSI PRODI TEKNIK INFORMATIKA IIB DARMAJAYA”** untuk memudahkan kelompok bidang keilmuan (KBK) melakukan deteksi tingkat kemiripan pada judul – judul yang diajukan yang ada pada jurusan teknik informatika dengan algoritma *TF-IDF* dan *Fuzzy Matching.*

## 1.2 Indentifikasi Masalah

Berdasarkan uraian pada latar belakang, ditemukan informasi mengenai sistem yang akan dibangun. Identifikasi masalah melalui latar belakang diantaranya:

1. Sulitnya bagi mahasiswa untuk mencari kemiripan ide judul mereka dengan judul yang sudah ada.
2. Pengecekan judul – judul mahasiswa yang diajukan masih dilakukan dengan cara manual yang tersimpan dalam bentuk *Microsoft Excel.*
3. Pertimbangan judul – judul yang telah diajukan ke kelompok bidang keilmuan (KBK) menjadi kurang efisien karena dilakukan satu persatu apabila judul yang diajukan sangat banyak, sehingga memungkinkan untuk terjadinya judul yang sama.

## 1.3 Rumusan Masalah

Perumusan masalah berdasarkan uraian latar belakang menjadi “Bagaimana kelompok bidang keilmuan (KBK) dapat mendeteksi kemiripan judul yang diajukan oleh mahasiswa teknik informatika dengan judul – judul yang sudah ada dan bagaimana mahasiswa dapat menentukan serta mendeteksi adanya kemiripan ide judulnya dengan judul yang sudah ada dengan menerapkan algoritma *TF-IDF* dan *Fuzzy Matching* untuk aplikasi berbasis android?”.

## 1.4 Batasan Masalah

Setiap penelitian memiliki batasan – batasan masalahnya. Batasan masalah untuk penelitian ini adalah:

1. Ruang lingkup penelitian ini hanya mencangkup program studi teknik informatika IIB Darmajaya.
2. Pengguna aplikasi ini adalah kelompok bidang keilmuan (KBK) dan mahasiswa program studi teknik informatika.
3. Kombinasi algoritma *TF-IDF* dan *Fuzzy Matching* memiliki kekurangan dalam memahami konteks judul – judul, data yang belum pernah ada, hubungan antar judul yang kompleks, serta sumber daya yang tinggi untuk melakukan komputasi dengan algoritma ini.
4. Menggunakan Bahasa pemrograman *Python* sebagai *backend computing* dan database *JSON* untuk penyimpanan data judul.

## 1.5 Tujuan Penelitian

Tujuan dari penelitian ini ialah:

1. Membantu menghindari adanya kemungkinan judul skripsi yang sama antara judul yang diajukan dengan judul yang sudah ada.
2. Meningkatkan efesiensi proses pertimbangan terhadap judul – judul yang diajukan oleh mahasiswa teknik informatika.
3. Membantu pendeteksian kemiripan judul yang diajukan secara efektif.

## 1.6 Manfaat Penelitian

Beberapa manfaat dari penelitian ini diantaranya ialah:

1. Membantu pengambilan keputusan pertimbangan kelompok bidang keilmuan (KBK) dalam menentukan judul – judul yang diajukan.
2. Membantu memberikan rekomendasi judul serupa yang cukup akurat.
3. Membantu dalam mendeteksi persentase kemiripan dari judul – judul yang diajukan.

## 1.7 Sistematika Penulisan

Sistematika penulisan yang digunakan dalam penyusunan penelitian ini adalah sebagai berikut:

**BAB I PENDAHULUAN**

Berisi penjelasan mengenai latar belakang identifikasi masalah, rumusan masalah, ruang lingkup penelitian, tujuan, manfaat dan sistematika penulisan.

**BAB II TINJAUAN PUSTAKA**

Berisi dasar teori yang digunakan sebagai acuan dalam pengambilan data, analisis, hasil penelitian dan pengujian.

**BAB III METODE PENELITIAN**

Dalam bab ini dijelaskan tentang metode pendekatan-pendekatan yang dilakukan untuk mendukung dalam proses penelitian.

**DAFTAR PUSTAKA**

# BAB II

# TINJAUAN PUSTAKA

## 2.1 Implementasi

Implementasi sebagai kegiatan mendistribusikan keluaran dari suatu kebijakan yang dijalankan oleh seorang pelaksana (untuk menyampaikan keluaran kebijakan) kepada suatu kelompok sasaran dalam upaya mencapai kebijakan tersebut [3].

## 2.2 Algoritma

Algoritma adalah sekumpulan instruksi atau langkah-langkah yang dituliskan secara sistematis dan digunakan untuk menyelesaikan suatu persoalan atau suatu permasalahan matematika dan logika dengan bantuan komputer [4].

Algoritma sebagai suatu upaya dengan urutan operasi yang disusun secara logis dan sistematis untuk menyelesaikan suatu masalah untuk menghasilkan suatu output tertentu [5].

Dari beberapa pengertian sebelumnya, dapat disimpulkan bahwa algoritma adalah serangkaian langkah-langkah yang terdefinisi dengan baik yang digunakan untuk menyelesaikan suatu masalah. Algoritma haruslah jelas, terstruktur, dan terdefinisi dengan baik agar dapat dimengerti dan diterapkan dengan benar.

## 2.3 Deteksi

Kata “deteksi” berasal dari bahasa latin *”Detectio”* yang berarti penemuan atau pengungkapan, dalam Bahasa Indonesia, deteksi merupakan penemuan, pengenalan, atau penentuan terhadap suatu objek[6].

## 2.4 Aplikasi Android

Aplikasi Android adalah satu unit perangkat lunak yang dibuat untuk melayani kebutuhan akan beberapa aktivitas seperti sistem perniagaan, game, pelayanan masyarakat, periklanan, atau semua proses yang hampir dilakukan manusia [7], sedangkan menurut Nazaruddin, Aplikasi Android adalah perangkat lunak yang berjalan pada platform Android. Platform Android adalah sistem operasi mobile yang dikembangkan oleh Google dan bersifat open source.

Aplikasi Android adalah program komputer yang dirancang untuk dijalankan pada perangkat mobile yang menggunakan sistem operasi Android. Aplikasi Android dapat digunakan untuk berbagai macam keperluan, seperti komunikasi, hiburan, edukasi, dan bisnis[8]. Sedangkan menurut Sibero Aplikasi Android adalah program yang dibuat dengan menggunakan bahasa pemrograman Java dan SDK (Software Development Kit) Android. Aplikasi Android dapat diunduh dan dipasang pada perangkat Android dari Google Play Store atau sumber pihak ketiga lainnya.

Sehingga dapan disimpulkan bahwa aplikasi Android adalah program komputer yang dirancang untuk dijalankan pada perangkat mobile yang menggunakan sistem operasi Android. Aplikasi Android dapat digunakan untuk berbagai macam keperluan, dan dapat diunduh dan dipasang pada perangkat Android dari berbagai sumber.

Beberapa ciri – ciri aplikasi android diantaranya:

* Dibuat dengan menggunakan bahasa pemrograman Java dan SDK Android.
* Dapat diunduh dan dipasang pada perangkat Android dari Google Play Store atau sumber pihak ketiga lainnya.
* Berjalan pada sistem operasi Android.
* Digunakan untuk berbagai macam keperluan, seperti komunikasi, hiburan, edukasi, dan bisnis.

## 2.5 Algoritma *TF-IDF*

Algoritma *Term Frequency–Inverse Document Frequency* merupkan metode yang digunakan untuk menghitung bobot atau nilai setiap kata kunci dalam sebuah dokumen. Bobot atau nilai ini mencerminkan seberapa penting kata kunci dalam dokumen dan seberapa jarang kata kunci muncul di dalam dokumen.

Algoritma TF-IDF banyak digunakan dalam berbagai aplikasi pengolahan teks, seperti:

* Pencarian informasi: Algoritma ini membantu mesin pencari menemukan dokumen yang relevan dengan kueri pengguna.
* Klasifikasi teks: Algoritma ini membantu mengklasifikasikan dokumen ke dalam kategori tertentu.
* Ringkasan teks: Algoritma ini membantu mengidentifikasi kata kunci penting dalam dokumen dan membuat ringkasan yang informatif.
* Analisis sentimen: Algoritma ini membantu menentukan opini atau perasaan yang terkandung dalam sebuah teks.

### 2.5.1 Term Frequency

*Term Frequency* merupakan metode yang bekerja dengan cara mengukur seberapa sering kata kunci muncul dalam dokumen. Semakin sering kata kunci muncul, semakin tinggi nilai TF-nya.

*tf(t, d) =*

Keterangan:

tf(*t, d*) = Nilai kata yang sering muncul dalam kumpulan dokumen.

*t* = *term*

*d* = dokumen

### 2.5.2 Inverse Document Frequency

Sedangkan *Inverse Document Frequency*,Mengukur seberapa jarang kata kunci muncul di seluruh dokumen dalam kumpulan data. Semakin jarang kata kunci muncul, semakin tinggi nilai IDF-nya.

*idf(t, D) =*

Keterangan:

idf(*t, D*) = Nilai kata yang jarang muncul dalam kumpulan dokumen.

*t* = *term*

N = Total nilai dokumen dalam sebuah list.

*d* = Dokumen

D = Total dokumen

Nilai *TF-IDF* dihitung dengan mengalikan nilai *TF* dan nilai *IDF*, sehingga rumusnya adalah:

*tfidf(t, d, D) = tf(t, d) \* idf(t, D)*

Keterangan:

tf(*t, d*) = Total nilai kata yang sering muncul dalam kumpulan dokumen

idf(*t, D*) = Nilai kata yang jarang muncul dalam kumpulan dokumen.

*t* = *term*

*d* = Dokumen

D = Total dokumen

### 2.5.3 Langkah Langkah Algoritma *TF – IDF*

Misalkan kita memiliki dua dokumen:

* Dokumen 1: "Mobil berwarna merah itu melaju kencang di jalan tol."
* Dokumen 2: "Sepeda motor adalah kendaraan roda dua yang banyak digunakan di Indonesia."

Mari kita hitung nilai TF-IDF untuk kata kunci "mobil" pada kedua dokumen:

* Dokumen 1:

TF: Kata kunci "mobil" muncul 1 kali.

IDF: Asumsikan kata kunci "mobil" muncul dalam 10% dokumen dalam kumpulan data. Maka, IDF = log(N/n), di mana N adalah jumlah total dokumen dan n adalah jumlah dokumen yang mengandung kata kunci "mobil". Dalam kasus ini, IDF = log(10/1) = 2.3025.

TF-IDF = 1 \* 2.3025 = 2.3025

* Dokumen 2:

TF: Kata kunci "mobil" tidak muncul.

IDF: IDF sama seperti pada Dokumen 1.

TF-IDF = 0 \* 2.3025 = 0

Dari contoh ini, kita dapat melihat bahwa nilai TF-IDF untuk kata kunci "mobil" pada Dokumen 1 lebih tinggi daripada Dokumen 2. Hal ini menunjukkan bahwa kata kunci "mobil" lebih penting bagi Dokumen 1 dibandingkan Dokumen 2.

## 2.6 Algoritma Fuzzy Matching

Algoritma Fuzzy Matching adalah teknik pencocokan string yang mempertimbangkan kemiripan antara dua string, bukan hanya persamaan kedua string. Hal ini memungkinkan algoritma untuk menemukan kecocokan bahkan ketika terdapat perbedaan kecil dalam ejaan, tata bahasa, atau struktur kalimat.

Algoritma Fuzzy Matching bekerja dengan menghitung jarak antara dua string. Jarak ini mewakili jumlah operasi yang diperlukan untuk mengubah satu string menjadi string lainnya. Operasi ini dapat berupa:

* Penyisipan: Menambahkan karakter ke string
* Penghapusan: Menghapus karakter dari string
* Penggantian: Mengubah karakter dalam string

Setelah menghitung jarak string, algoritma Fuzzy Matching akan menentukan apakah dua string cukup mirip untuk dianggap sebagai kecocokan. Nilai ambang batas kesamaan biasanya digunakan untuk menentukan hal ini.

### 2.6.1 Cara Kerja Algoritma Fuzzy Matching

Berikut adalah contoh bagaimana algoritma Fuzzy Matching dapat digunakan untuk menemukan kecocokan antara dua string:

String 1: "Sepatu lari"

String 2: "Sepatu lari merah"

Algoritma Levenshtein Distance akan menghitung jarak antara dua string ini sebagai satu. Hal ini karena hanya perlu satu operasi (penghapusan kata "merah") untuk mengubah String 1 menjadi String 2. Misalkan nilai ambang batas kesamaan untuk algoritma Fuzzy Matching adalah 0.8. Dalam hal ini, algoritma akan menganggap String 1 dan String 2 sebagai kecocokan karena jaraknya (0.8) lebih besar dari nilai ambang batas.

## 2.7 UML (Unified Modeling Language)

UML adalah salah satu standar bahasa yang banyak digunakan di dunia industri untuk mendefinisikan requirement, membuat analisis dan desain, serta menggambarkan arsitektur dalam pemrograman berorientasi objek [9].

Perancangan Sistem Terdapat beberapa bagian dalam perancangan UML (Unified Modelling Language) yaitu sebagai berikut:

### 2.7.1 Activity Diagram

*Activity diagram* merupakan alur kerja dari sebuah sistem yang digambarkan melalui diagram [10].

## 2.8 Database MySQL()

*MySQL* adalah *database* *multiuser* yang menggunakan bahasa *Structured* *Query Language* (*SQL*) dan termasuk dalam kategori *Database* *Management System* (*DBMS*) yang bersifat *open* *source* [11]. *MySQL* menyimpan data dalam tabel yang terstruktur. Setiap tabel terdiri dari baris dan kolom. Baris mewakili catatan individual, sedangkan kolom mewakili atribut atau karakteristik dari catatan tersebut. Misalnya, tabel "Pelanggan" mungkin memiliki kolom untuk nama, alamat, dan nomor telepon pelanggan. *MySQL* menggunakan bahasa *query* terstruktur (*SQL*) untuk mengakses dan memanipulasi data*.*

## 2.9 Pengujian *Black Box*

Pengujian *Black Box* merupakan pengujian perangkat lunak dimana pengujian berasal dari persyaratan fungsional tanpa menguji desain dan kode program. Pengujian dilakukan untuk mengetahui setiap fungsi, masukan, dan keluaran dari program sesuai dengan persyaratan yang dibutuhkan. Pengujian *black box* dilakukan dengan membuat kasus – kasus yang bersifat menguji semua fungsi dengan alat uji apakah sesuai dengan spesifikasi yang dibutuhkan.

## 2.10 Hasil Penelitian Sebelumnya

Berikut ini adalah beberapa literatur yang digunakan dalam penelitian, dapat dilihat pada **Tabel 2.10.1** :

**Tabel 2.10.1** Tinjauan Pustaka Jurnal

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **No** | **Nama, Tahun** | **Judul** | **Tujuan Penelitian** | **Hasil Penelitian** |
| 1 | Herlina Sari, 2021 | Penerapan Algoritma Text Mining Dan TF-IDF Untuk Pengelompokan Topik Skripsi Pada Aplikasi Repository STMIK Budi Darma | Tujuan dari penelitian yang dibahas di halaman ini adalah untuk mengembangkan aplikasi repository yang dapat mengelompokkan topik skripsi secara otomatis menggunakan algoritma Text Mining dan TF-IDF. | Studi ini menggali bagaimana teknik text mining dan algoritma TF-IDF dapat digunakan untuk mengklasifikasikan dan mengelompokkan topik skripsi yang ada dalam repositori digital. |
| 2 | Ni Putu Anggi Yuliani, 2017 | Sistem Temu Balik Dokumen Teks Menggunakan Metodebooleandan *Term Wight Tf.Idf* | Tujuan penelitian dari jurnal tersebut adalah untuk mengembangkan Sistem Temu Balik Dokumen Teks yang efisien menggunakan metode Boolean dan Term Weight TF.IDF. | Metode Boolean: Penelitian ini menggunakan metode Boolean untuk memproses pencarian dokumen teks.  Term Weight TF.IDF: Penelitian ini juga mengimplementasikan pembobotan term menggunakan TF.IDF untuk meningkatkan akurasi hasil temu balik.  Efisiensi Sistem: Hasil penelitian menunjukkan peningkatan efisiensi dalam sistem temu balik dokumen teks.  Implementasi: Penelitian ini memberikan wawasan tentang implementasi sistem yang diusulkan dalam praktik nyata. |
| 3 | Nilawati | Penggunaan Metode Cosine Similarity Dan Tf-Idf Untuk Klasifikasi Judul Seminar Proposal Pada Fakultas Teknik Universitas Jabal Ghafur | Tujuan penelitian ini untuk mengklasifikasikan judul seminar proposal pada Fakultas Teknik Universitas Jabal Ghafur dengan metode Cosine Similarity dan TF-IDF, berfokus pada pengembangan sistem secara otomatis mengkategorikan judul berdasarkan kesamaan konten, sehingga memmbantu pengelolaan pengetahuan dan organisasi informasi seminar. | Penelitian ini menghasilkan sebuah sistem klasifikasi judul seminar proposal yang efektif. Sistem ini menggunakan metode Cosine Similarity dan TF-IDF (Term Frequency-Inverse Document Frequency) untuk mengukur kesamaan antara dokumen teks. |
| 4 | Dhebys Suryani Hormansyah, 2018 | Penerapan Metode Tf-Idf Dan N-Gram Pada Pengembangan Aplikasi Chatbot Berbasis Line Untuk Layanan Publik Kesehatan Di Kota Malang | Tujuan dari penelitian ini adalah untuk mengembangkan aplikasi chatbot berbasis LINE yang menggunakan metode TF-IDF dan N-GRAM untuk meningkatkan layanan publik kesehatan di Kota Malang. | Penelitian tersebut mengembangkan aplikasi chatbot berbasis LINE yang menggunakan metode TF-IDF dan N-gram untuk meningkatkan layanan publik kesehatan di Kota Malang. |
| 5 | Kairunnisa, 2019 | Penerapan Fuzzy String Matching Pada Aplikasi Pencarian Judul Skripsi Teknik Informatika Berbasis Android | Tujuan dari penelitian ini adalah untuk mengembangkan sebuah aplikasi repository berbasis Android yang memanfaatkan algoritma Fuzzy String Matching | Aplikasi ini dirancang untuk membantu mahasiswa dalam mencari referensi skripsi dengan cepat dan efisien. Dengan aplikasi ini, pengguna dapat melakukan pencarian judul skripsi pada STMIK Budi Darma dan mendapatkan saran kata yang relevan, |

# BAB III METODOLOGI PENELITIAN

## 3.1 Metode Pengumpulan Data

Dalam metode pengumpulan data memiliki peranan yang penting dalam mendapatkan suatu informasi. Data yang relevan dengan pokok pembahasannya adalah indikator keberhasilan suatu penelitian. Dalam pengumpulan data, beberapa metode yang digunakan diantaranya:

### 3.1.1 Wawancara

Wawancara adalah proses pengumpulan data dengan tanya jawab secara langsung dengan orang – orang yang terkait. Berikut ini merupakan beberapa pertanyaan wawancara yang dilakukan yaitu:

1. Kendala apa saja yang dialami oleh kelompok bidang keilmuan (KBK)?
2. Berapa lama periode mahasiswa mengumpulkan judul skripsi?

### 3.1.2 Pengamatan

Penulis mengamati secara langsung lingkungan sistem yang akan dibangun. Ini bertujuan untuk mendapatkan gambaran yang jelas terhadap sistem yang akan dibangun.

### 3.1.3 Tinjauan Pustaka

Dalam pengumpulan data pada prosesnya dilakukan dengan mempelajari berbagai bentuk bahan tertulis yang berupa buku – buku, artikel, dokumen – dokumen, yang terkait secara langsung dengan sistem kemiripan judul skripsi.

## 3.2 Implementasi Algoritma *TF-IDF* Dan *Fuzzy Matching*

Dalam penelitian ini, untuk melakukan kombinasi algoritma *TF-IDF* dengan algoritma *Fuzzy Matching* terhadap dataset judul-judul skripsi Teknik Informatika yang berbentuk *spreadsheet* dapat dilakukan dengan langkah-langkah sebagai berikut.

### 3.2.1 Pra-Proses *Dataset*

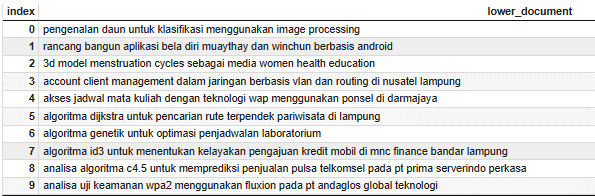
Langkah pertama yang harus dilakukan ialah pra-proses *dataset*, yang dapat dilakukan dengan cara menghapus data duplikat, menghapus data yang tidak relavan, mengubah huruf-hurufnya menjadi kecil, dan menghapus spasi pada awal dan akhir kalimat. Data redudan dan data tidak valid dapat dilihat pada **Tabel 3.2.1.1.**

**Tabel 3.2.1.1** Data Redudan dan Data tidak Valid



Pada **Tabel 3.2.1.1**, merupakan dataset yang sudah di hapus data-data yang tidak relavan, dan menghapus data duplikat.

**Tabel 3.2.1.2** Merubah Huruf-Huruf Menjadi Kecil

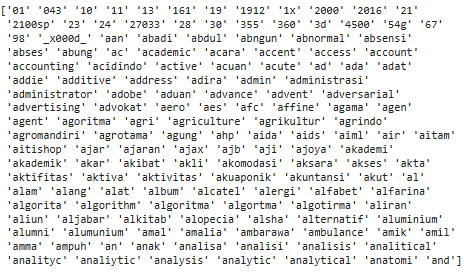


Dalam **Tabel 3.2.1.2**, merupakan hasil mengubah huruf-huruf pada dataset menjadi huruf kecil (*lowercase*) sehingga memudahkan proses kalkulasi pada algoritma *TF-IDF* dan *Fuzzy Matching*.

### 3.2.2 Menghitung Nilai Kemiripan Dengan Algoritma *TF-IDF*

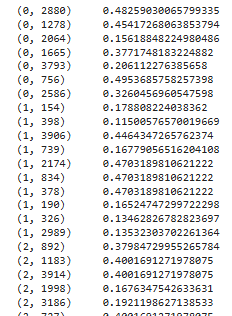
Setelah di proses datasetnya selanjutnya adalah menghitung nilai kemiripannya. Algoritma pertama yang dilakukan dalam perhitungan ini ialah algoritma *TF-IDF* (*Term Frequency-Inverse Document Frequency*). Untuk *query* yang akan digunakan adalah “*perancangan sistem pakar untuk penyakit katarak*”.

Sebelum melakukan perhitungan dengan algoritma *TF-IDF* hal pertama yang harus dilakukan adalah mengubah dokumen ke dalam bentuk vector dengan melakukan tokenisasi.



**Gambar 3.2.2.1** Tokenisasi dokumen

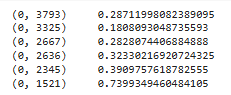
Pada **Gambar 3.2.2.1**, terlihat proses tokenisasi pada dokumen. Tokenisasi merupakan proses mengubah, memecah urutan teks menjadi unit-unit yang lebih kecil yang disebut token. Token ini dapat berupa kata, frasa, simbol, atau elemen bermakna lainnya [12].



**Gambar 3.2.2.2** Proses Vektorisasi

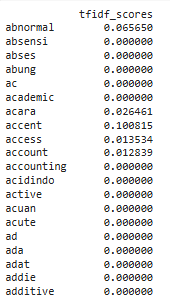
Pada **Gambar 3.2.2.2**, merupakan proses vektorisasi dari proses tokenisasi. Hal ini dilakukan agar kata atau kalimat dapat diproses dalam algoritma *TF-IDF* yang memerlukan vektor angka-angka.

Langkah selanjutnya adalah melakukan *vectorize* terhadap *query* yang dimana *query*-nya adalah “*perancangan sistem pakar untuk penyakit katarak*”.



**Gambar 3.2.2.3** Proses Vektorisasi *Query*

Pada **Gambar 3.2.2.2**, dapat terlihat hasil vektorisasi terhadap *query* atau *input*. Selanjutnya setelah didapatkan nilai vektor dari *query* atau *input,* dikalikan dengan vektor *document* yang dibalik atau *Transpose*. *Transpose* merupakan teknik mengubah kolom menjadi baris dan baris menjadi kolom [13].

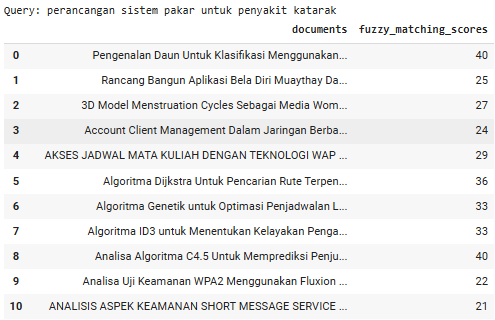


**Gambar 3.2.2.4** Perkalian Matrix *Query* Dengan *Document*

Pada **Gambar 3.2.2.4** terlihat hasil dari perkalian *matrix* antara matriks *query* dengan matriks *document* yang menentukan nilai *TF-IDF* nya.

### 3.2.3 Menghitung Nilai Kemiripan Dengan Algoritma *Fuzzy Mathcing*

Langkah selanjutnya setelah didapatkan nilai *TF-IDF* nya adalah menghitung nilai *query* terhadap *document* dengan algoritma *Fuzzy Matching* yang nantinya akan dikalikan dengan algoritma *TF-IDF.*



**Tabel 3.2.3.1** Nilai *Query* Terhadap Document Dengan Algoritma *Fuzzy Matching*

Pada **Tabel 3.2.3.1** dapat dilihat hasil perhitungan jarak antara *query* dengan setiap *text* di dalam *document.* Langkah terakhir untuk mengkombinasikan kedua algoritma *TF-IDF* dan *Fuzzy Matching* ialah dengan mengkalikan hasil dari kedua algoritma tersebut.

### 3.2.4 Kombinasi Algoritma *TF-IDF* dan *Fuzzy Matching*

Setelah didapatkan nilai dari *TF-IDF* dengan *Fuzzy Matching,* langkah selanjutnya adalah mengkombinasikan hasil kedua algoritma tersebut dengan mengkalikan nilainya.



**Gambar 3.2.4.1** Kombinasi Nilai *TF-IDF* dan *Fuzzy Matching*

Pada **Gambar 3.2.4.1** terlihat hasil dari kombinasi kedua algoritma, dimana kombinasi dilakukan dengan perkalian sederhana berdasarkan nilai *TF-IDF* dengan *Fuzzy Matching*.

## 3.3 Analisis Sistem

Dalam perancangan sebuah sistem, diperlukan langkah untuk mengidentifikasi dan mengevaluasi permasalahan atau hambatan dalam proses perancangan. Adapun kebutuhan dalam perancangan yang akan diimplementasikan diantaranya, perangkat keras dan perangkat lunak.

### 3.2.1 Analisis Kebutuhan Perangkat Lunak

Dalam implementasi aplikasi perangkat lunak, maka software yang dibutuhkan adalah sebagai berikut:

1. Sistem Operasi Windows
2. Bahasa Pemrograman Java
3. Android Studio
4. *Internet*

### 3.3 Perancangan Sistem

Desain aplikasi dibuat untuk memudahkan dalam pembuatan aplikasi yang berjalan berdasarkan kebutuhan sistem pada alur sistem proses pengecekan kemiripan judul skripsi.



**Gambar 3.3.1** *Flowchart* Proses Pengeckan

Berdasarkan alur pada **Gambar 3.3.1** proses pengecekan kemiripan judul skripsi, dimulai dari memasukkan judul skripsi, selanjutnya sistem akan memproses tingkat kemiripan judul skripsi yang dimasukkan dengan beberapa judul skripsi yang sudah pernah ada. Apabila tingkat kemiripan judul skripsi tinggi, maka akan ditolak terhadap judul yang diajukan oleh mahasiswa, dan apabila tingkat kemiripan tidak terlalu tinggi maka judul memiliki kemungkinan untuk diterima.

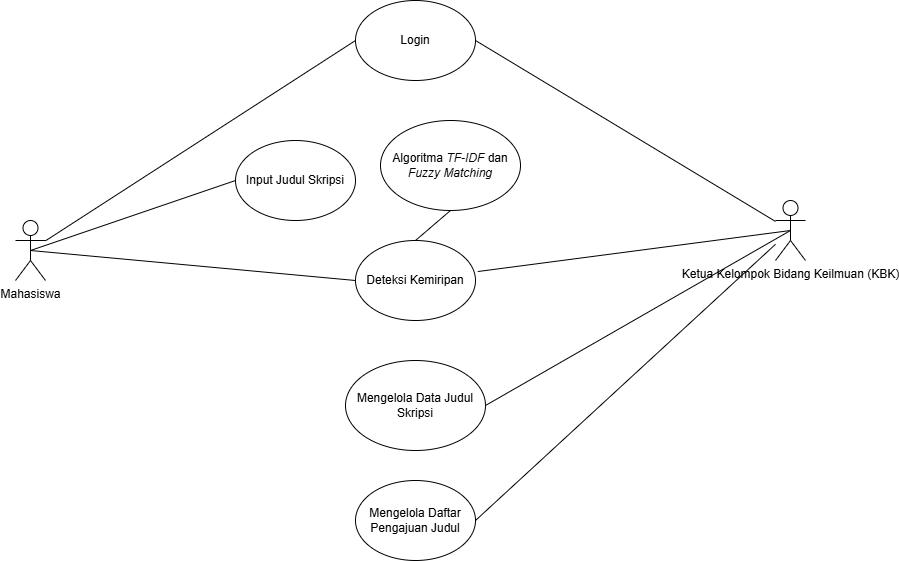
Berdasarkan masalah dan alur sistem yang ada, maka akan dibangun sistem yang dapat mengukur tingkat kemiripan judul skripsi menggunakan algoritma *TF-IDF* dengan kombinasi algoritma *Fuzzy Matching*.

## 3.4 Tahapan Membangun *Prototype*

*Tools* yang digunakan dalam perancangan ini adalah UML yaitu penggambaran *use case diagram, activity diagram,* dan *class diagram.*

### 3.4.1 Penggambaran *Usecase Diagram*

*Usecase Diagram* merupakan *tool* yang mendeskripsikan interaksi antar satu atau lebih aktor dengan sistem informasi yang akan dibuat. Berikut merupakan *usecase diagram* dari sistem deteksi kemiripan judul skripsi.



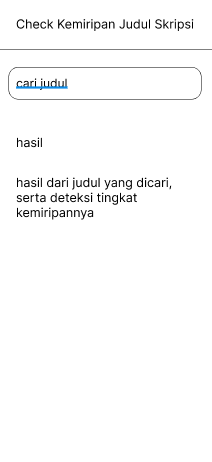
**Gambar 3.4.1** Diagram *Usecase*

## 3.5 Desain Aplikasi

Desain aplikasi melibatkan pembuatan antarmuka pengguna yang memfasilitasi pengguna dalam mencapai tujuan mereka dengan cara yang efektif dan efisien [14].

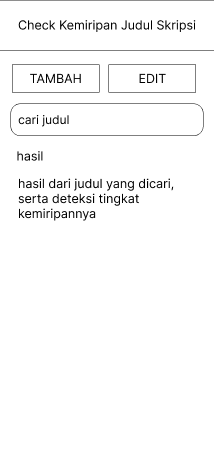
### 3.5.1 Desain Tampilan Awal Aplikasi

Tampilan awal merupakan tampilan yang dihasilkan oleh program saat pertama kali program dijalankan. *prototype* dari tampilan awal aplikasi deteksi kemiripan judul skripsi dapat terlihat pada **Gambar 3.5.1**.

****

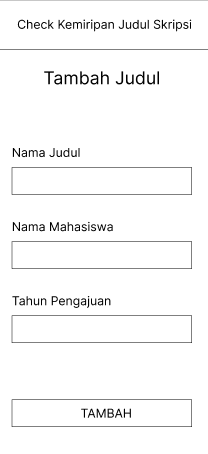
**Gambar 3.5.1** *Prototype* Tampilan Awal Aplikasi

**3.5.3 Tampilan Awal Hak Akses Ketua KBK**

****

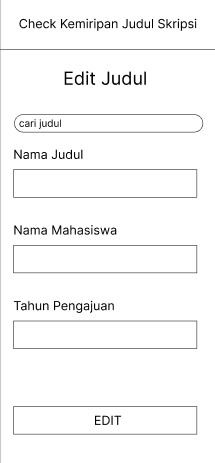
**Gambar 3.5.3** *Prototype* Tampilan Awal Admin

### 3.5.4 Tampilan Tambah Judul Hak Akses Ketua KBK

****

**Gambar 3.5.4** *Prototype* Tampilan Tambah Judul

### 3.5.5 Tampilan Edit Judul Hak Akses Ketua KBK

****

**Gambar 3.5.5** *Prototype* Tampilan Edit Judul

## 3.6 Jadwal Penelitian

**Tabel 3.6.1** Jadwal Penelitian

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| No | Kegiatan | Bulan Ke-1 | | | | Bulan Ke-2 | | | |
| 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 |
| 1 | Wawancara Dan Pengumpulan Data |  |  |  |  |  |  |  |  |
| 2 | Merancang dan Mendesain Sistem |  |  |  |  |  |  |  |  |
| 3 | Mengembangan Sistem |  |  |  |  |  |  |  |  |
| 4 | Implementasi Sistem |  |  |  |  |  |  |  |  |
| 5 | Testing dan Pengujian Sistem |  |  |  |  |  |  |  |  |

# BAB IV

# HASIL DAN PEMBAHASAN

## 4.1 Hasil Penelitian

Penelitian ini menghasilkan penggunaan kombinasi algoritma TDF-IDF dan Fuzzy Matching yang diimplementasikan kedalam bentuk aplikasi **Deteksi Kemiripan Judul Skripsi Berbasis Android**.

## 4.1.1 Algoritma *TF-IDF* dan *Fuzzy Matching*

Pada hasil penelitian ini, masing-masing algoritma memiliki kelebihan dan kekurangannya dalam mendeteksi kemiripan dua atau lebih teks, sehingga peneliti melakukan kombinasi dari kedua algoritma ini untuk menghasilkan hasil kecocokan teks yang lebih akurat.

**Algoritma *TF-IDF***

Untuk mengukur tingkat

2

Pada hasil penelitian ini, peneliti melakukan kombinasi algoritma *TF-IDF* dan *Fuzzy Matching* dengan Bahasa Pemrograman *Python* yang selanjutnya diterapkan kedalam bentuk aplikasi *Android*. Terdapat 2 aplikasi yang dibuat pada penelitian ini yaitu aplikasi *Android* untuk digunakan oleh siapa saja yang ingin mendeteksi kemiripan judul skripsinya dengan judul skripsi yang sudah pernah diajukan, dan aplikasi berbasis *Website* untuk Ketua Kelompok Bidang Keilmuan (KBK) meng*input*kan judul skripsi terbaru atau mengelola judul skripsi.

## 4.2 Kombinasi Algoritma *TF-IDF* Dengan *Fuzzy Matching*

## 4.3 Implementasi Sistem *Restful API*

Pada langkah ini, peneliti mengimplementasikan kombinasi algoritma *TF-IDF* dan *Fuzzy Matching* ke dalam bentuk *Restful API* dengan menggunakan bahasa pemrograman *Python* yang nantinya akan digunakan oleh aplikasi *Android* untuk dipakai mendeteksi kemiripan judul skripsi, dan aplikasi *Website* yang digunakan oleh Ketua Kelompok Bidang Keilmuan (KBK) dalam mengelola judul skripsi.

### 4.3.1 …

## 4.4 Aplikasi Android

Selanjutnya setelah mengaplikasikan sistem untuk mendeteksi kemiripan judul skripsi yaitu, memanggil *Restful API* dengan mengimplementasikannya pada aplikasi *Android.* Berikut merupakan tampilan halaman utama dari aplikasi “Cek Judul Skripsi”

### 4.1.2 Tampilan Halaman Cek Judul Skripsi

Berikut merupakan tampilan halaman cek judul skripsi.

### 4.1.3 Tampilan Halaman Tanya *Artificial Intelligence*

# DAFTAR PUSTAKA

[1] Y. Gu, Y. Wang, J. Huan, Y. Sun, and S. Xu, “An improved TFIDF algorithm based on dual parallel adaptive computing model,” *Int. J. Embed. Syst.*, 2020, doi: 10.1504/IJES.2020.108278.

[2] L. Mo, “Fuzzy matching algorithm of network information retrieval based on discrete mathematics,” *Appl. Nanosci.*, 2023, doi: 10.1007/s13204-021-02190-y.

[3] M. A. Dalma, “Implementasi : Pengertian, Tujuan dan Contoh,” *Dosen Pintar*. 2021.

[4] Sutiono, “Pengertian Algoritma, Flowchart dan Pseudocode,” *DosenIT.com*. 2017.

[5] R. Abidin, “Pengertian Algoritma Pemrograman,” 25 April.

[6] R. E. Izzaty, B. Astuti, and N. Cholimah, “Pengertian dan Tujuan Deteksi Deteksi,” *Angew. Chemie Int. Ed. 6(11), 951–952.*, 2019.

[7] I. A. Fahmy, “5 Pengertian Aplikasi Android Menurut Para Ahli,” Ani Ramdhani.

[8] Suparyanto dan Rosad, “Pengertian Aplikasi Mobile Dan Android,” *Suparyanto dan Rosad*, 2020.

[9] R. A. S and M. Salahudin, “Rekayasa Perangkat Lunak Terstruktur dan Berorientasi Objek,” in *Bandung: Informatika*, 2013.

[10] A. Paramitha, “Materi 4 - activity diagram,” *Mater. 4 - Act. Diagr. APSI - 2*, 2018.

[11] D. Agustini and M. Farida, “SISTEM INFORMASI PENGELOLAAN PASAR CEMARA BANJARMASIN BERBASIS WEB,” *Technol. J. Ilm.*, 2021, doi: 10.31602/tji.v12i1.4174.

[12] R. Sennrich, B. Haddow, and A. Birch, “Neural machine translation of rare words with subword units,” in *54th Annual Meeting of the Association for Computational Linguistics, ACL 2016 - Long Papers*, 2016. doi: 10.18653/v1/p16-1162.

[13] A. Radul, A. Paszke, R. Frostig, M. J. Johnson, and D. Maclaurin, “You only Linearize Once: Tangents Transpose to Gradients,” *Proc. ACM Program. Lang.*, 2023, doi: 10.1145/3571236.

[14] A. Cooper, R. Reimann, D. Cronin, and C. Noessel, *About Face: The Essentials of Interaction Design, 4th Edition*. 2014.