



Kombinasi Algoritma *TF-IDF* dan *Fuzzy Matching* untuk Deteksi Kemiripan Judul Skripsi

Muhammad Fauzan Azima¹, Arif Nur Listanto^{*2}, Fitria³, Chairani⁴

^{1, *2, 3, 4}Program Studi Teknik Informatika, Institut Informatika Dan Bisnis Darmajaya, Lampung, Indonesia

*Email Penulis Korespondensi: anurlistanto@gmail.com

Abstrak

Dalam era informasi digital, deteksi kemiripan teks menjadi penting untuk berbagai aplikasi seperti plagiarisme, pengelompokan dokumen, dan penyaringan informasi. Penelitian ini bertujuan untuk mengembangkan metode yang efektif dalam mendeteksi kemiripan teks dengan menggabungkan algoritma TF-IDF dan Fuzzy Matching. Alasan pemilihan topik ini didasarkan pada kebutuhan akan akurasi yang lebih tinggi dalam mengidentifikasi kemiripan teks yang seringkali tidak dapat diatasi dengan metode konvensional secara memadai. Metode penelitian ini melibatkan penggunaan TF-IDF untuk mengekstraksi fitur penting dari teks, yang kemudian dipadukan dengan Fuzzy Matching untuk mengatasi variasi dan ketidakpastian dalam teks yang dibandingkan. Hasil penelitian menunjukkan bahwa kombinasi kedua algoritma ini mampu meningkatkan akurasi deteksi kemiripan teks dibandingkan dengan penggunaan salah satu algoritma secara terpisah. Pengujian dilakukan pada dataset Judul Skripsi Prodi Teknik Informatika IIB Darmajaya dengan variasi teks, dan hasilnya menunjukkan peningkatan yang signifikan dalam tingkat keakuratan yang tinggi. Evaluasi model menggunakan precision menunjukan akurasi sebesar 88,89%. Kesimpulan penelitian ini menegaskan pentingnya pendekatan gabungan TF-IDF dan Fuzzy Matching dalam aplikasi deteksi kemiripan teks, yang dapat memberikan kontribusi signifikan terhadap peningkatan kualitas dan efisiensi pengelolaan informasi digital.

Kata kunci— Kemiripan Teks, TF-IDF, Fuzzy Matching, Deteksi Plagiarisme

Abstract

In the era of digital information, text similarity detection becomes important for various applications such as plagiarism, document clustering, and information filtering. This research aims to develop an effective method in detecting text similarity by combining TF-IDF and Fuzzy Matching algorithms. The reason for choosing this topic is based on the need for higher accuracy in identifying text similarity which often cannot be adequately addressed by conventional methods. The research method involves the use of TF-IDF to extract important features from the text, which is then combined with Fuzzy Matching to address variations and uncertainties in the compared texts. The results show that the combination of these two algorithms is able to improve the accuracy of text similarity detection compared to the use of either algorithm separately. Tests were conducted on the IIB Darmajaya Informatics Study Program Thesis Title dataset with text

variations, and the results showed a significant increase in the high level of accuracy. The evaluation of the model using precision showed an accuracy of 88.,89%. The conclusion of this research confirms the importance of the combined approach of TF-IDF and Fuzzy Matching in text similarity detection applications, which can make a significant contribution to improving the quality and efficiency of digital information management.

Keywords— Text Similarity, TF-IDF, Fuzzy Matching, Plagiarism Detection

1. PENDAHULUAN

Dalam era digital yang semakin berkembang, pemrosesan teks telah menjadi bidang yang sangat penting, terutama dalam pendeteksian *plagiarisme*, pencarian informasi, dan analisis *sentiment*. Dalam suatu instansi akademik, penelitian atau biasa dikenal dengan istilah tugas akhir bagi mahasiswa merupakan sks wajib yang tersaji pada KRS setiap program studi atau jurusan di seluruh perguruan tinggi [1].

Setiap judul skripsi harus unik yang berarti judul skripsi tidak boleh ada kesamaan baik berupa judul maupun konten yang ada didalamnya harus bersifat orisinal. Hal tersebut menjadi syarat diterimanya pengajuan judul skripsi. Setiap judul skripsi yang diajukan oleh mahasiswa akan di filter oleh ketua Kelompok Bidang Keilmuan (KBK). Di Darmajaya khususnya prodi Teknik Informatika terdapat 3 konsentersasi bidang keilmuan yaitu *Artificial Intelligence*, *Mobile Technology*, dan *Multimedia*. Masalah yang sering dijumpai yaitu judul yang diajukan oleh mahasiswa seringkali mirip, sangat mirip, dan bahkan identik dengan judul yang sudah pernah diajukan sebelumnya, ataupun mahasiswa hanya mengganti judul dengan lokasi penelitian yang berbeda tanpa adanya nilai keterbaharuan dari penelitian tersebut.

Penelitian sebelumnya membahas tentang penggunaan algoritma *Winnowing* untuk mendeteksi *plagiarisme* dalam judul skripsi. Algoritma ini mencari kesamaan kata atau kalimat dengan membandingkan *'fingerprint'* dari teks. Penelitian ini menggunakan teknik *hashing*, algoritma mengubah *n-gram* teks menjadi nilai *hash* yang kemudian dibandingkan untuk menemukan kesamaan [2]. Namun pada penelitian ini, melakukan kombinasi algoritma *TF-IDF* dan *Fuzzy Matching* untuk membuat sistem deteksi kemiripan judul skripsi.

Penelitian selanjutnya yaitu implementasi satu algoritma untuk mendeteksi kemiripan judul skripsi menggunakan Algoritma *Smith Waterman*. Dengan algoritma tersebut didapatkan persentase kemiripan judul skripsi pada Teknik Informatika FIKOM UMI untuk tahun 2018-2019. Penelitian ini mendeteksi kecocokan judul skripsi dengan presentase kemiripan tertinggi antara judul yang sama [3]. Dalam penelitian ini, peneliti menggunakan kombinasi *TF-IDF* dengan *Fuzzy Matching* sehingga memberikan hasil yang lebih komprehensif dalam mendeteksi kemiripan judul skripsi.

Penelitian berikutnya yang dilakukan oleh Fitrianiingsih [4] dalam pengembangan sistem deteksi kemiripan judul skripsi dengan menggunakan algoritma *Oliver* yang merupakan fungsi di dalam bahasa pemrograman *PHP* untuk menghasilkan nilai kemiripan antara dua *string*. Dimana sistem ini dirancang untuk mengatasi masalah redudansi judul skripsi dengan membandingkan judul yang diajukan oleh mahasiswa terhadap *database* judul skripsi yang ada. Sistem ini bekerja dengan menggunakan data uji dan data *training*, yang mampu menolak secara otomatis pada judul dengan kemiripan diatas 60%.

Penelitian ini sangat penting untuk dilakukan dalam rangka mengantisipasi judul skripsi mahasiswa yang dicurigai terdapat tindak *plagiarisme* sehingga mahasiswa dapat melakukan filter terlebih dahulu sebelum mengajukan judul tersebut ke prodi, selain itu proses ini akan membantu ketua Kelompok Bidang Keilmuan (KBK) dalam mempertimbangkan rekomendasi judul skripsi yang diajukan oleh mahasiswa apakah terdapat kemiripan atau tidak. Adapun penelitian ini menggabungkan 2 buah algoritma yaitu *TF-IDF* dan *Fuzzy Matching* agar dapat mendeteksi berapa tingkat kemiripan judul yang diajukan dibandingkan dengan judul-judul yang sudah ada sebelumnya.

2. METODE PENELITIAN

2.1. Teori Algoritma TF-IDF

TF-IDF adalah sebuah metode yang merupakan integrasi antar *term frequency (TF)*, dan *inverse document frequency (IDF)*. *Term Frequency* dihitung menggunakan Persamaan dengan *term frequency* ke-*i* adalah frekuensi kemunculan *term* ke-*i* dalam dokumen ke-*j*. *Inverse Document Frequency (IDF)* adalah logaritma dari rasio jumlah seluruh dokumen dalam korpus dengan jumlah dokumen yang memiliki *term* yang dimaksud seperti yang dituliskan secara matematis pada Persamaan [5][6]. Algoritma *TF-IDF* dalam rumusnya dituliskan seperti pada formula 1, formula 2, dan formula 3.

Pada *TF (Term Frequency)* rumusnya adalah sebagai berikut:

$$tf(t, d) = \frac{f_{t,d}}{\sum_{t' \in d} f_{t',d}} \quad (1)$$

dimana:

t : *term*
d : *document*

Sedangkan *IDF (Inverse Document Frequency)* rumusnya adalah sebagai berikut:

$$idf(t, D) = \log \frac{N}{|\{d: d \in D \text{ and } t \in d\}|} \quad (2)$$

dimana:

t : *term*
d : *document*
D : *total document*
N : *number of document*

Dan untuk rumus keseluruna dari *TF-IDF* sebagai berikut:

$$tfidf(t, d, D) = tf(t, d) * idf(t, D) \quad (3)$$

2.2. Teori Algoritma Fuzzy Matching

Fuzzy Matching adalah teknik yang digunakan untuk mengidentifikasi tingkat kesamaan antara *string*, yang sangat berguna untuk penyelarasan ontologi di mana pencocokan yang tepat tidak selalu memungkinkan karena adanya variasi data. Makalah ini membahas beberapa algoritma untuk pencocokan *string fuzzy*, termasuk *Jaro-Winkler* dan *Levenshtein* [7][8][9]. Dalam penggunaan algoritma *Fuzzy Matching* terdapat beberapa langkah diantaranya:

1. Preprocessing:

Langkah ini dilakukan untuk mengubah huruf menjadi kecil, menghindari masalah dengan *case sensitive*.

2. Menghitung Similarity Score

Dalam langkah ini, merupakan metode untuk menghitung seberapa mirip dua string, beberapa metode yang umum digunakan diantaranya *Levenshtein Distance*, *Jaro-Winkler Distance* dan *Cosine Similarity* [10].

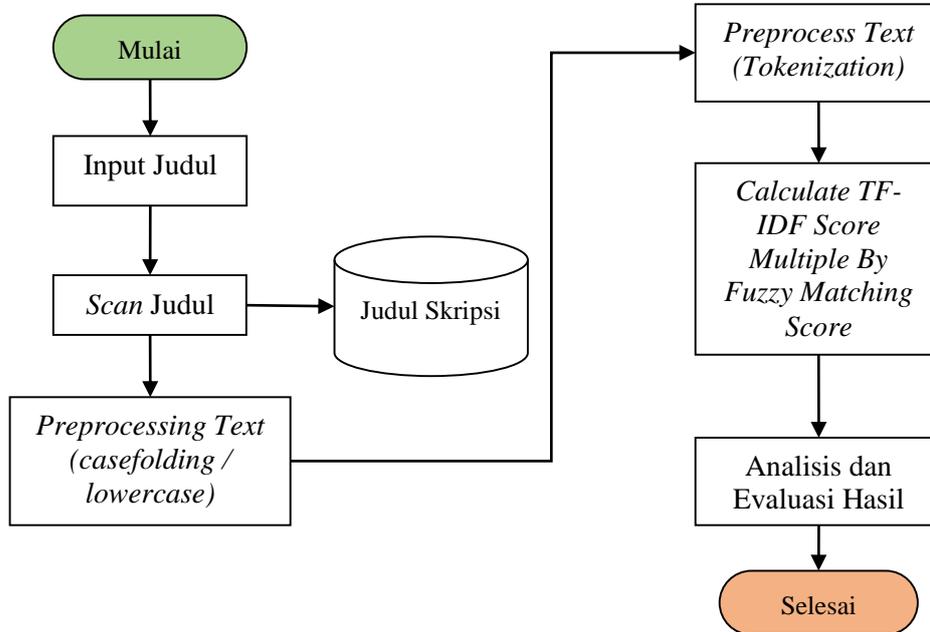
3. Membandingkan String

Langkah ini membandingkan *string* yang sudah dihitung *similarity score* pada langkah sebelumnya dan menemukan atau memilih nilai *similarity score* tertinggi dalam pencocokan *string*.

2.3. Alur Penelitian

Dalam proses penelitian ini merupakan implementasi *Flowchart* dari kombinasi algoritma *TF-IDF (Term Frequency-Inverse Document Frequency)* untuk mendeteksi kemiripan

judul skripsi dan menggambarkan proses-proses secara jelas di dalam sistem. *Flowchart* digunakan secara luas dalam pengembangan *software*, perancangan teknik, dan penelitian ilmiah [11]. Pada perancangan sistem ini juga membahas implementasi algoritma *TF-IDF* dan *Fuzzy Matching* dengan bahasa pemrograman *Python*. Alur (*Flowchart*) dari sistem pendeteksi kemiripan judul skripsi seperti dapat dilihat pada Gambar 1.



Gambar 1 Alur Penelitian (*Flowchart*) Sistem

Berdasarkan alur penelitian pada Gambar 1, terlihat bahwa pada langkah pertama pengguna diharuskan menginputkan judul yang akan dilakukan deteksi kemiripan dengan mengetikkan secara manual pada *input* yang tersedia di dalam sistem.

index	lower_document
2168	Rancang Bangun Model Pembelajaran Pengenalan Hardware Komputer Berbasis Augmented Reality Pada SMK Budi Karya Natar
2169	Perangkat Lunak Sistem Perombaan (Rekayasa Perangkat Lunak, Jaringan Komputer, dan Multimedia) Di Prodi Teknik Informatika Tingkat SMA / SMK Sederajat IIB Darmajaya Berbasis Web
2170	Penerapan Teknologi Augmented Reality Pada Edukasi Rambu Rambu Lalu Lintas Berbahasa Inggris Untuk Siswa Sekolah Dasar Menggunakan Metode Marker Based Tracking
2171	Pengembangan Media Pembelajaran Mata kuliah Keamanan Komputer dan Jaringan Berbasis Film Animasi 2 Dimensi
2172	Implementasi Metode Image Stitching Untuk Visualisasi 360 Derajat Kampus IIB Darmajaya Berbasis Website
2173	Penerapan Multimedia Interaktif Pada Mata Kuliah Teori Bahasa Otomata Menggunakan Teknik Animasi 2 Dimensi
2174	Sistem Prediksi Masa Studi Mahasiswa Menggunakan Metode KNN
2175	Implementasi Game Edukasi Pengenalan Pahlawan Nasional Berbasis Android Untuk Siswa SMP Kelas 8 Menggunakan Algoritma Shuffle Random
2176	Deteksi Detak Jantung Dengan Sensor dan Kamera Berbasis Android
2177	Visualisasi Data Akademik Mahasiswa Program Studi Teknik Informatika IIB Darmajaya Berbasis Web Menggunakan Metode Bubble Sort

Gambar 2 Sampel Data Akhir Judul Skripsi

Tahap pertama yang dilakukan adalah memanggil semua data judul skripsi di dalam *database* ke dalam sistem, dapat terlihat pada Gambar 1. Langkah selanjutnya yaitu mengubah semua huruf menjadi huruf kecil atau *casefolding (lowercase)* yang ditunjukkan pada Gambar 2.

index	lower_document
2168	rancang bangun model pembelajaran pengenalan hardware komputer berbasis augmented reality pada smk budi karya natar
2169	perangkat lunak sistem perlombaan (rekayasa perangkat lunak, jaringan komputer, dan multimedia) di prodi teknik informatika tingkat sma / smk sederajat iib darmajaya berbasis web
2170	penerapan teknologi augmented reality pada edukasi rambu rambu lalu lintas berbahasa inggris untuk siswa sekolah dasar menggunakan metode marker based tracking
2171	pengembangan media pembelajaran mata kuliah keamanan komputer dan jaringan berbasis film animasi 2 dimensi
2172	implementasi metode image stitching untuk visualisasi 360 derajat kampus iib darmajaya berbasis website
2173	penerapan multimedia interaktif pada mata kuliah teori bahasa otomata menggunakan teknik animasi 2 dimensi
2174	sistem prediksi masa studi mahasiswa menggunakan metode knn
2175	implementasi game edukasi pengenalan pahlawan nasional berbasis android untuk siswa smp kelas 8 menggunakan algoritma shuffle random
2176	deteksi detak jantung dengan sensor dan kamera berbasis android
2177	visualisasi data akademik mahasiswa program studi teknik informatika iib darmajaya berbasis web menggunakan metode bubble sort

Gambar 3 Hasil *CaseFolding (LowerCase)*

Setelah data dilakukan proses *casefolding (lowercase)* yang dapat terlihat pada Gambar 3, langkah selanjutnya yaitu dilakukan proses *tokenization* yang dapat terlihat pada Gambar 4.

```
['01' '043' '10' '11' '13' '161' '19' '1912' '1x' '2000' '2016' '21'
'2100sp' '23' '24' '27033' '28' '30' '355' '360' '3d' '4500' '54g' '67'
'98' '_x000d_' 'aan' 'abadi' 'abdul' 'abngun' 'abnormal' 'absensi'
'abses' 'abung' 'ac' 'academic' 'acara' 'accent' 'access' 'account'
'accounting' 'acidindo' 'active' 'acuan' 'acute' 'ad' 'ada' 'adat'
'addie' 'additive' 'address' 'adira' 'admin' 'administrasi'
'administrator' 'adobe' 'aduan' 'advance' 'advent' 'adversarial'
'advertising' 'advokat' 'aero' 'aes' 'afc' 'affine' 'agama' 'agen'
'agent' 'agoritma' 'agri' 'agriculture' 'agrikultur' 'agrindo'
'agromandiri' 'agrotama' 'agung' 'ahp' 'aida' 'aids' 'aiml' 'air' 'aitam'
'aitishop' 'ajar' 'ajaran' 'ajax' 'ajb' 'aji' 'ajoya' 'akademi'
'akademik' 'akar' 'akibat' 'akli' 'akomodasi' 'aksara' 'akses' 'akta'
'aktifitas' 'aktiva' 'aktivitas' 'akuaponik' 'akuntansi' 'akut' 'al'
'alam' 'alang' 'alat' 'album' 'alcatel' 'alergi' 'alfabet' 'alfarina'
'algoritma' 'algorithm' 'algoritma' 'algortma' 'algotirma' 'aliran'
'aliun' 'aljabar' 'alkitab' 'alopecia' 'alsha' 'alternatif' 'aluminium'
'alumni' 'alumunium' 'amal' 'amalia' 'ambarawa' 'ambulance' 'amik' 'amil'
'amma' 'ampuh' 'an' 'anak' 'analisa' 'analisi' 'analisis' 'analitical'
'analityc' 'analiytic' 'analysis' 'analytic' 'analytical' 'anatomi' 'and']
```

Gambar 4 Data hasil proses *Tokenization* 150 Token

Langkah selanjutnya setelah proses *tokenization* adalah menghitung nilai *TF-IDF* dan dikalikan dengan nilai *Fuzzy Matching* terhadap judul yang *diinputkan* oleh pengguna. Rumusnya dalam melakukan perhitungan kombinasi algoritma *TF-IDF* dengan *Fuzzy Matching* adalah sebagai berikut:

$$F(x) = \sum_t \text{Similar}(t) \text{TF} - \text{IDF}(t, d, D) \quad (4)$$

Score atau nilai yang dihasilkan berupa bilangan bulat atau bilangan desimal. Sebagai contoh kata “kucing”, dengan kata “kulit” memiliki *similarity score* sekitar 0,55, apabila dikalikan dengan 100, maka akan mendapatkan nilai persentase sekitar 55%.

Untuk melakukan pengukuran evaluasi model, dalam penelitian ini menggunakan evaluasi *precision*. Evaluasi *precision* merupakan metode ukur yang digunakan dalam pengukuran pencarian informasi, *machine learning*, dan *data science* dengan cara mengukur keakuratan prediksi positif yang dibuat oleh model [12]. *Precision* dapat didefinisikan sebagai rasio prediksi positif yang benar terhadap jumlah positif benar dengan positif salah. Dalam matematika *precision* dapat dirumuskan seperti pada formula (5).

$$\text{precision} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Positives (FP)}} \quad (5)$$

3. HASIL DAN PEMBAHASAN

Pembahasan terhadap hasil dan pengujian yang diperoleh dari penelitian ini adalah sebagai berikut:

3.1 Dataset Judul Skripsi

Dalam penelitian ini, peneliti menggunakan *dataset* Judul Skripsi Program Studi Teknik Informatika, IIB Darmajaya, dengan jumlah sekitar ± 2.148 judul skripsi pada *spreadsheet*. *Dataset* dapat terlihat pada Tabel 1.

Tabel 1 *Dataset* Judul Skripsi Prodi Teknik Informatika

NO	NPM	Nama	JudulSkripsi
1	09010090	Muhammad Alfath	Pengenalan Daun Untuk Klasifikasi Menggunakan Image Processing
2	097010044	Eko Budi Wicaksono	AKSES JADWAL MATA KULIAH DENGAN TEKNOLOGI WAP MENGGUNAKAN PONSEL DI DARMAJAYA
3	1211010008	Nur Rachmanto	Account Client Management Dalam Jaringan Berbasis VLAN dan Routing Di NUSATEL Lampung
4	1211010092	Febri Kurniatama	Rancang Bangun Aplikasi Bela Diri Muaythay Dan Winchun Berbasis Android
5	1311010112	Dwi Nanda Widiatama	Analisa Uji Keamanan WPA2 Menggunakan Fluxion Pada PT Andaglos Global Teknologi
6	1411010015	Selvi Novitasari	Algoritma ID3 untuk Menentukan Kelayakan Pengajuan Kredit Mobil di MNC Finance Bandar Lampung
7	1411010111	Widia Savitri	Algoritma Dijkstra Untuk Pencarian Rute Terpendek Pariwisata Di Lampung
8	1711010009	Alfreda Budi Mahatma Gultom	Analisa Algoritma C4.5 Untuk Memprediksi Penjualan Pulsa TELKOMSEL Pada PT Prima Serverindo Perkasa
9	1711010059	RICHO VERNANDO	3D Model Menstruation Cycles Sebagai Media Women Health Education
10	1711010074	Fatzrul ZN	Algoritma Genetik untuk Optimasi Penjadwalan Laboratorium
dst..

3.2. Proses Perhitungan Similarity Score Pada Input Terhadap Data

Untuk menghitung nilai “*similarity score*”, pada penelitian ini dilakukan perhitungan nilai *TF-IDF* terlebih dahulu setelahnya *Fuzzy Matching* dimana hasil dari kedua algoritma itu dapat dikombinasikan. Hasil dari perhitungan nilai *TF-IDF* terhadap *dataset* dengan input “*perancangan sistem pakar untuk penyakit katarak*”

	tfidf_scores
abnormal	0.042684
absensi	0.000000
abses	0.000000
abung	0.077099
ac	0.000000
academic	0.018631
acara	0.095337
accent	0.020596
access	0.014454
account	0.074099
accounting	0.051236
acidindo	0.000000
active	0.072677
acuan	0.000000
acute	0.187871
ad	0.019089
ada	0.015870
adat	0.000000
addie	0.071099
additive	0.020341

Gambar 5 Hasil *TF-IDF* Score

Dapat terlihat pada Gambar 5, untuk input “*perancangan sistem pakar untuk penyakit katarak*” menghasilkan *score* seperti pada tabel 3 terhadap beberapa *vector TF-IDF*. Langkah selanjutnya yaitu menghitung nilai *Fuzzy Matching* pada Gambar 6 input “*perancangan sistem pakar untuk penyakit katarak*”.

	fuzzy_score
abnormal	0.44
absensi	0.08
abses	0.25
abung	0.55
ac	0.29
academic	0.33
acara	0.68
accent	0.32
access	0.23
account	0.55
accounting	0.45
acidindo	0.30
active	0.48
acuan	0.31
acute	0.73
ad	0.42
ada	0.40
adat	0.34
addie	0.30
additive	0.36

Gambar 6 Hasil *Fuzzy Matching*

Berdasarkan Gambar 5 dan Gambar 6, didapatkan *score* untuk masing-masing algoritma, sehingga langkah terakhir ialah menggabungkan kedua algoritma tersebut dengan cari dikalikan *matrix score* dari kedua algoritma tersebut, maka didapatkanlah hasil seperti berikut.

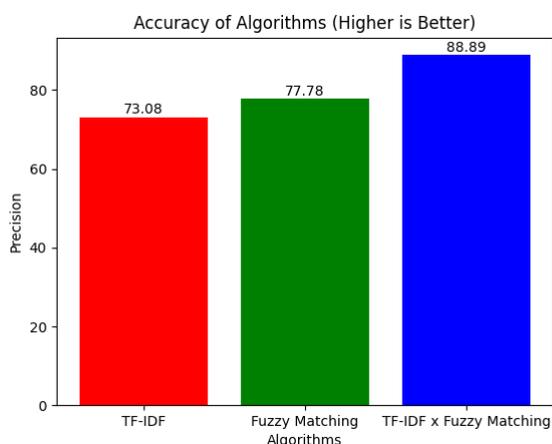
```
0.0),
('PERANCANGAN SISTEM PENDETEKSI GERAKAN OBJEK KAMERA UNTUK PEMANTAUAN DAN KEAMANAN RUANGAN.',
0.07387110615925219),
('PERANCANGAN SISTEM INFORMASI AKADEMIK SMK AL-FAJAR KEC. KASUI KABUPATEN WAY KANAN',
0.0322343890430008),
('PERANCANGAN WEBSITE SEBAGAI SARANA PROMOSI PADA PT. ASURANSI TAKAFUL KELUARGA CABANG LAMPUNG MENGGUNAKAN ASP',
0.016935274069585032),
('SISTEM KOMPUTERISASI SIMPAN PINJAM PADA KOPERASI SAUDARA 28 PADANG CERMIN',
0.005297353504006998),
('DESAIN WEB ORGANISASI KEMAHASISWAAN (MPM, BEM, UKM, HIMA) PERGURUAN TINGGI DARMAJAYA',
0.0),
('KOMPUTERISASI PENGOLAHAN DATA PENJUALAN BARANG PADA PT. LAUTAN LUAS Tbk BANDAR LAMPUNG BERBASIS MULTIUSER MENGGUNAKAN VISUAL BASIC 6.0 ',
0.0),
('SISTEM INFORMASI PENJUALAN SEMEN PADA PT. SEMEN BATURAJA (PERSERO) PANJANG BANDAR LAMPUNG DENGAN MENGGUNAKAN APLIKASI VISUAL BASIC 6.0',
0.0048140080860779935),
('SISTEM IKLAN BARIS BERBASIS WEB PADA SKH LAMPUNG EKSPRES PLUS',
0.0063161951561487685).
```

Gambar 7 Hasil dari *score similarity* terhadap *dataset*

Berdasarkan pada Gambar 7, didapatkan nilai / *score* dengan *input* yang dimasukkan oleh *user* terhadap *dataset* Judul Skripsi Prodi Teknik Informatika IIB Darmajaya.

3.3. Evaluasi Algoritma TF-IDF dan Fuzzy Matching

Dalam penelitian ini, dilakukan evaluasi terhadap *model* kombinasi algoritma TF-IDF dengan Fuzzy Matching, serta evaluasi akurasi terhadap masing-masing algoritma TF-IDF dan Fuzzy Matching yang dapat dilihat pada Gambar 8.



Gambar 8 Perbandingan Akurasi Algoritma

Pada Gambar 8, dapat dilihat bahwa skor akurasi dengan menggunakan metode *precision* dari algoritma *TF-IDF* sendiri dan algoritma *Fuzzy Matching* sendiri berada dibawah 80%, sedangkan apabila kedua algoritma tersebut digabungkan, maka menghasilkan tingkat akurasi yang tinggi 88,89% atau 90%. Selanjutnya pada penelitian ini dilakukan pengujian model kombinasi algoritma *TF-IDF* dengan *Fuzzy Matching* sebanyak 10 kali percobaan yang dapat terlihat pada Gambar 9.

index	Judul	Test Judul	Tingkat Kemiripan
0	Pengenalan Daun Untuk Klasifikasi Menggunakan Image Processing	Klasifikasi Daun Menggunakan Teknik Pemrosesan Citra	65.27%
1	Rancang Bangun Aplikasi Bela Diri Muaythay Dan Winchun Berbasis Android	Pengembangan Aplikasi Android untuk Pembelajaran Seni Bela Diri	39.12%
2	3D Model Menstruation Cycles Sebagai Media Women Health Education	Pembuatan Model 2D Siklus Menstruasi untuk Edukasi Kesehatan Wanita	32.15%
3	Account Client Management Dalam Jaringan Berbasis VLAN dan Routing Di NUSATEL Lampung	Implementasi Manajemen Akun Klien Menggunakan VLAN dan Routing pada Jaringan	46.68%
4	AKSES JADWAL MATA KULIAH DENGAN TEKNOLOGI WAP MENGGUNAKAN PONSEL DI DARMAJAYA	Sistem Informasi Akademik Berbasis Web untuk Akses Jadwal Kuliah	22.42%
5	Algoritma Dijkstra Untuk Pencarian Rute Terpendek Pariwisata Di Lampung	Implementasi Algoritma A* dalam Pencarian Jalur Terpendek Wisata Lampung	69.08%
6	Algoritma Genetik untuk Optimasi Penjadwalan Laboratorium	Penerapan Algoritma Particle Swarm Optimization untuk Penjadwalan Lab	75.75%
7	Algoritma ID3 untuk Menentukan Kelayakan Pengajuan Kredit Mobil di MNC Finance Bandar Lampung	Penggunaan Decision Tree dalam Analisis Kelayakan Kredit Kendaraan	34.6%
8	Analisa Algoritma C4.5 Untuk Memprediksi Penjualan Pulsa TELKOMSEL Pada PT Prima Serverindo Perkasa	Implementasi Naive Bayes untuk Prediksi Penjualan Produk Telekomunikasi	33.14%
9	Analisa Uji Keamanan WPA2 Menggunakan Fluxion Pada PT Andaglos Global Teknologi	Evaluasi Keamanan Protokol WPA3 dengan Metode Penetration Testing	25.42%
10	ANALISIS ASPEK KEAMANAN SHORT MESSAGE SERVICE (SMS) PADA JALUR KOMUNIKASI GSM	Studi Perbandingan Keamanan Komunikasi SMS dan Instant Messaging	26.2%

Gambar 9 Hasil percobaan model kombinasi algoritma *TF-IDF* dengan *Fuzzy Matching*

Pada pengujian model, yang dapat terlihat pada Gambar 9. Model algoritma kombinasi *TF-IDF* dengan *Fuzzy Matching* mampu memberikan tingkat persentase kemiripan judul yang akurat, baik gaya penulisan judul hingga keunikan kata di dalam judul.

3.4 Kelebihan dan Kekurangan Kombinasi Algoritma *TF-IDF* Dengan *Fuzzy Matching*

Dalam penelitian ini, kombinasi algoritma *TF-IDF* dan *Fuzzy Matching* mampu untuk mendeteksi tingkat kemiripan judul yang memiliki beberapa kelebihan diantaranya akurasi yang tinggi, fleksibilitas terhadap variasi dan ketidakpastian dalam teks yang dibandingkan serta implementasi yang praktis dalam pengembangan aplikasi *Android*. Namun penelitian ini juga memiliki beberapa kekurangan diantaranya membutuhkan sumber daya komputasi yang tinggi apabila digunakan dalam data dengan jumlah yang sangat besar, dan keterbatasan data pada dataset tertentu (judul skripsi dari program studi Teknik Informatika) yang tidak sepenuhnya representatif untuk judul skripsi dari berbagai domain.

3.5. Merancang Aplikasi *Android*

Dalam penelitian ini, peneliti merancang hasil dari kombinasi algoritma *TF-IDF* dan *Fuzzy Matching* dalam bentuk aplikasi *Android* [13], agar dapat digunakan untuk melakukan proses pengecekan judul skripsi oleh mahasiswa, serta merancang sistem *chatting* untuk mahasiswa dapat bertukar pikiran dengan *Artificial Intelligence*, sehingga membantu mahasiswa dalam menemukan dan menentukan judul ataupun tema yang akan diangkat di dalam skripsinya.

Pada Gambar 10, tampak tampilan halaman menu pada aplikasi. Terdapat halaman “Cek Judul” untuk mengecek judul apakah memiliki tingkat kemiripan yang tinggi dengan judul yang ada. Selanjutnya terdapat menu “Tanya Judul” di dalam menu ini, pengguna atau *user* dapat bertanya dengan *Artificial Intelligence* seputar judul-judul skripsi teknik informatika, dan memberikan beberapa ide judul skripsi untuk pengguna atau *user*.



Gambar 10 Tampilan Menu Aplikasi



Gambar 11 Hasil Cek Kemiripan Judul Skripsi

Pada Gambar 11, merupakan tampilan dari hasil pengecekan judul skripsi dengan input "perancangan sistem pakar untuk penyakit katarak" yang dimana memiliki tingkat kemiripan yang tinggi terhadap judul skripsi "Perancangan sistem pakar untuk mendiagnosa penyakit katarak" dengan persentase 89,89%.



Gambar 12 Halaman *Chatting Dengan Artificial Intelligence*

Pada Gambar 12, peneliti menambahkan fitur untuk dapat bertukar pikiran atau *Brainstorm* dengan teknologi *Artificial Intelligence*. Sehingga mahasiswa yang sedang atau ingin mengambil skripsi tetapi tidak tahu harus mulai dari mana dapat menggunakan fitur ini untuk menemukan ide dan menentukan judul skripsinya.

4. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa metode yang dikembangkan melalui kombinasi algoritma *TF-IDF* dan *Fuzzy Matching* dalam deteksi kemiripan teks menunjukkan hasil yang signifikan. Penelitian ini fokus pada penerapan metode tersebut dalam konteks judul skripsi, di mana kebutuhan akan akurasi tinggi sangat penting untuk menghindari *plagiarisme* dan memastikan orisinalitas. Dengan memanfaatkan kekuatan dari kedua algoritma ini, penelitian ini berhasil mengatasi beberapa tantangan utama yang sering muncul dalam deteksi kemiripan teks.

5. SARAN

Untuk penelitian lebih lanjut yang dapat mengatasi beberapa kekurangan dalam penelitian ini, peneliti dapat memperluas *dataset* yang digunakan dalam pengujian agar mencakup judul skripsi dari berbagai institusi dan disiplin ilmu, sehingga hasil penelitian dapat lebih representatif dan dapat digeneralisasikan. Selain itu, pengembangan teknik optimasi untuk mengurangi kompleksitas perhitungan dari kombinasi algoritma *TF-IDF* dan *Fuzzy Matching* dapat diterapkan dengan lebih efisien dan cepat pada berbagai *platform* dengan sumber daya komputasi yang bervariasi [14]. Penelitian selanjutnya juga dapat memfokuskan pada peningkatan akurasi melalui pengintegrasian metode ini dengan teknologi kecerdasan buatan lainnya, seperti *machine learning* atau *deep learning* [15], untuk memberikan hasil deteksi yang lebih presisi dan mampu menangani variasi teks yang lebih luas.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Tim Redaksi Jurnal Teknik Politeknik Negeri Sriwijaya yang telah memberi memberi kesempatan, sehingga artikel ilmiah ini dapat diterbitkan.

DAFTAR PUSTAKA

- [1] Mauliddin, "Analisis Kesalahan Penggunaan Uji Statistik pada Skripsi Mahasiswa," *J. Mat. dan Pembelajaran*, 2017.
- [2] N. Alamsyah, "DETEKSI PLAGIARISME TINGKAT KEMIRIPAN JUDUL SKRIPSI DENGAN ALGORITMA WINNOWER," *Technol. J. Ilm.*, 2017, doi: 10.31602/tji.v8i4.1119.
- [3] U. N. Hasanah, R. Satra, and F. Umar, "Deteksi Kemiripan Judul Skripsi Menggunakan Algoritma Smith Waterman," *Bul. Sist. Inf. dan Teknol. Islam*, 2020, doi: 10.33096/busiti.v1i1.676.
- [4] N. Fitriyaningsih, M. Asfi, D. Prasetyo, R. P. Kusuma, and M. A. Sulhan, "Deteksi Tingkat Kemiripan Judul Menggunakan Algoritma Oliver Pada Sistem Informasi Pengajuan Skripsi," *J. MEDIA Inform. BUDIDARMA*, 2022, doi: 10.30865/mib.v6i4.4409.
- [5] 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.
- [6] N. K. Widyasanti, I. K. G. Darma Putra, and N. K. Dwi Rusjyanthi, "Seleksi Fitur Bobot Kata dengan Metode TFIDF untuk Ringkasan Bahasa Indonesia," *J. Ilm. Merpati (Menara Penelit. Akad. Teknol. Informasi)*, 2018, doi: 10.24843/jim.2018.v06.i02.p06.
- [7] T. Efrianto and M. Hayaty, "JARO WINKLER ALGORITHM FOR MEASURING SIMILARITY ONLINE NEWS," *J. Tek. Inform.*, 2022.
- [8] B. Berger, M. S. Waterman, and Y. W. Yu, "Levenshtein Distance, Sequence Comparison and Biological Database Search," *IEEE Trans. Inf. Theory*, 2021, doi: 10.1109/TIT.2020.2996543.
- [9] B. Bhasuran, G. Murugesan, S. Abdulkadhar, and J. Natarajan, "Stacked ensemble combined with fuzzy matching for biomedical named entity recognition of diseases," *J. Biomed. Inform.*, 2016, doi: 10.1016/j.jbi.2016.09.009.
- [10] F. Rahutomo, T. Kitasuka, and M. Aritsugi, "Semantic cosine similarity," *7th Int. student Conf. Adv. Sci. Technol. ICAST*, 2012.
- [11] P. Zhang, W. Dou, and H. Liu, "Hierarchical data structures for flowchart," *Sci. Rep.*, 2023, doi: 10.1038/s41598-023-31968-z.
- [12] G. Indrayanto, "Application of Accuracy and Precision Evaluations Based on the Current United States and Indonesian Pharmacopoeias: A Critical Review," *Makara J. Sci.*, 2022, doi: 10.7454/mss.v26i4.1343.
- [13] S. N. L. Muhammad Fauzan Azima, "Rancang Bangun Aplikasi Kamus Bahasa dan Aksara Lampung Dialek A dan Dialek O Berbasis Android," *J. Tek.*, 2020.
- [14] H. J. Kim, J. W. Baik, and K. Chung, "Optimization of associative knowledge graph using TF-IDF based ranking score," *Appl. Sci.*, 2020, doi: 10.3390/app10134590.
- [15] L. Chen, P. Chen, and Z. Lin, "Artificial Intelligence in Education: A Review," *IEEE Access*, 2020, doi: 10.1109/ACCESS.2020.2988510.