



Similarity Identification Model of Thesis Titles with Mahalanobis Distance Approach

Muhammad Fajri Munawar¹, Rudi Heriansyah^{2*}, Muhammad Haviz Irfani³, Muhammad Ikhwan Jambak⁴, Dwi Asa Ferano⁵

^{1,2,3,4,5}Teknik Informatika, Ilmu Komputer dan Sains, Universitas Indo Global Mandiri²Jurusan, Fakultas, Institusi

¹mfajrimunawar46@gmail.com, ²rudi@uigm.ac.id*, ³m.haviz@uigm.ac.id, ⁴jambak@uigm.ac.id, ⁵dwiasa@uigm.ac.id

ABSTRACT

This study aims to identify the similarity of thesis titles by applying the Mahalanobis Distance method which is known to be effective in measuring the distance between vectors by considering data distribution and correlation between variables. In its implementation, each thesis title is represented in vector form using the TF-IDF scheme before calculating the level of similarity using Mahalanobis Distance. The test results show that this method is able to produce similarity values between titles, but its performance has not shown optimal effectiveness in the context of similarity classification. The highest precision value obtained of 1.0 indicates that this method is quite reliable in identifying pairs of titles that are truly similar. However, the low recall value of only 0.5 indicates that there are many pairs of similar titles that fail to be detected, resulting in an F1-score value of only 0.638. This shows an imbalance between the system's ability to detect similarity and its classification accuracy. Although the accuracy value is relatively high, ranging from 0.958 to 0.988, these results do not necessarily reflect the overall effectiveness of the method in handling minor classification errors. Testing of the threshold parameters also shows that a value of 0.1 provides the best performance compared to other threshold values because it is able to maintain a balance between precision, recall, F1-score, and accuracy.

Keywords: Thesis, Mahalanobis Distance, identification, threshold, TF-IDF

ABSTRAK

Penelitian ini bertujuan untuk mengidentifikasi kemiripan judul skripsi dengan menerapkan metode Mahalanobis Distance yang dikenal efektif dalam mengukur jarak antar vektor dengan mempertimbangkan distribusi data dan korelasi antar variabel. Dalam implementasinya, setiap judul skripsi direpresentasikan dalam bentuk vektor menggunakan skema TF-IDF sebelum dihitung tingkat kemiripannya menggunakan *Mahalanobis Distance*. Hasil pengujian menunjukkan bahwa metode ini mampu menghasilkan nilai kemiripan antar judul, namun performanya belum menunjukkan efektivitas optimal dalam konteks klasifikasi kemiripan. Nilai presisi tertinggi yang diperoleh sebesar 1,0 menunjukkan bahwa metode ini cukup andal dalam mengidentifikasi pasangan judul yang benar-benar mirip. Namun demikian, rendahnya nilai recall yang hanya mencapai 0,5 mengindikasikan adanya banyak pasangan judul mirip yang gagal terdeteksi, sehingga menghasilkan nilai F1-score yang hanya mencapai 0,638. Hal ini menunjukkan ketidakseimbangan antara kemampuan sistem dalam mendeteksi kemiripan dan akurasi klasifikasinya. Meskipun nilai akurasi tergolong tinggi, berkisar antara 0,958 hingga 0,988, hasil tersebut tidak serta-merta mencerminkan efektivitas keseluruhan metode dalam menangani kesalahan klasifikasi minor. Pengujian terhadap parameter threshold juga menunjukkan bahwa nilai 0,1 memberikan performa terbaik dibandingkan dengan nilai threshold lainnya karena mampu menjaga keseimbangan antara precision, recall, F1-score, dan akurasi.

Kata kunci: Judul Skripsi, Mahalanobis Distance, identifikasi, *threshold*, TF-IDF

1. PENDAHULUAN

Pemilihan judul skripsi merupakan tahapan awal yang sangat penting bagi mahasiswa dalam proses penyusunan karya ilmiah. Judul yang baik harus memenuhi kriteria orisinalitas, relevansi, serta belum pernah digunakan dalam penelitian sebelumnya guna menghindari terjadinya duplikasi atau redundansi ilmiah[1]. Namun demikian, dalam

praktiknya masih sering ditemukan adanya kesamaan atau kemiripan antar judul skripsi, baik dari segi struktur kalimat maupun tema kajian. Hal ini menimbulkan permasalahan terkait keaslian karya ilmiah yang berpotensi menurunkan mutu akademik. Oleh sebab itu, diperlukan suatu mekanisme yang mampu membantu proses identifikasi kemiripan judul skripsi secara lebih sistematis dan mendalam.

Saat ini, sebagian besar perguruan tinggi telah memanfaatkan sistem katalog daring seperti *Online Public Access Catalog* (OPAC) untuk memfasilitasi pencarian referensi judul skripsi oleh mahasiswa. OPAC memungkinkan pencarian berdasarkan kata kunci atau frasa judul, namun sistem ini masih memiliki keterbatasan dalam hal pengukuran tingkat kemiripan antar judul[2]. Hal ini disebabkan oleh algoritma pencarian konvensional yang hanya mengandalkan pencocokan literal antar kata, tanpa mempertimbangkan aspek semantik, sinonim, maupun struktur kalimat yang lebih kompleks.

Kemampuan untuk mengukur tingkat kemiripan antar judul skripsi secara kuantitatif menjadi hal yang penting, tidak hanya bagi mahasiswa dalam memilih topik penelitian yang unik, tetapi juga bagi dosen pembimbing dan pengelola program studi dalam proses verifikasi judul. Dengan mengetahui tingkat kemiripan terhadap judul-judul yang telah ada, mahasiswa dapat menghindari topik yang terlalu mirip, sehingga mampu menghasilkan penelitian yang lebih orisinal dan bernilai ilmiah tinggi. Di sisi lain, sistem identifikasi semacam ini dapat meningkatkan efisiensi dan objektivitas dalam proses evaluasi dan persetujuan judul skripsi di lingkungan akademik[3].

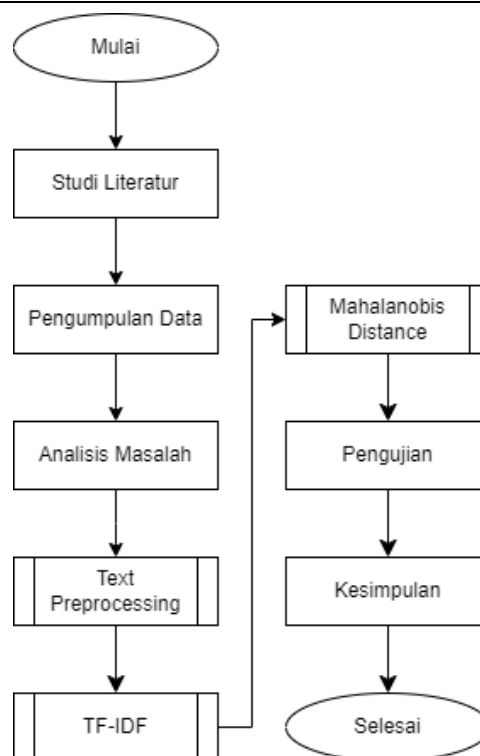
Salah satu pendekatan matematis yang dapat digunakan dalam pengukuran kemiripan teks adalah metode Mahalanobis Distance. Metode ini memiliki keunggulan dalam menghitung jarak antar vektor dengan mempertimbangkan korelasi antar variabel, sehingga lebih sensitif terhadap struktur data yang kompleks. Dalam konteks penelitian ini, judul skripsi direpresentasikan dalam bentuk vektor fitur, yang kemudian dianalisis menggunakan Mahalanobis Distance untuk menghitung tingkat kemiripannya [4]. Berbeda dengan metode jarak lainnya seperti Euclidean Distance, metode Mahalanobis mampu mengenali pola hubungan statistik antar kata atau frasa secara lebih akurat [5]. Selain itu, metode Mahalanobis dapat digunakan untuk menghitung data dengan jarak tertentu terhadap *mean* data dan metode Mahalanobis memiliki kelebihan yaitu dapat mengukur kemiripan dua buah objek dengan jumlah data yang berbeda[6].

Penelitian ini bertujuan untuk mengembangkan sistem identifikasi kemiripan judul skripsi berbasis metode Mahalanobis Distance. Diharapkan sistem yang dikembangkan dapat memberikan hasil analisis yang lebih akurat dalam mendeteksi kemiripan antar judul skripsi, serta menjadi solusi dalam mendukung upaya penjaminan mutu akademik, khususnya dalam menjamin orisinalitas karya ilmiah mahasiswa. Selain itu, hasil penelitian ini juga diharapkan dapat memberikan kontribusi dalam pengembangan sistem pendukung keputusan untuk pengelolaan tugas akhir di lingkungan perguruan tinggi.

2. METODE PENELITIAN

Tahapan dalam penelitian ini (Gambar 1) menunjukkan alur penelitian yang digunakan dalam proses identifikasi kemiripan judul skripsi menggunakan metode Mahalanobis Distance. Tahapan ini dimulai dari proses awal yaitu studi literatur, yang bertujuan untuk memahami konsep, teori, serta penelitian terdahulu yang relevan. Kemudian dilanjutkan dengan proses pengumpulan data berupa judul-judul skripsi yang akan dianalisis. Setelah data terkumpul, dilakukan analisis masalah untuk mengidentifikasi isu utama yang akan diselesaikan dalam penelitian ini.

Tahap berikutnya yaitu *text preprocessing*, yang mencakup proses pembersihan teks seperti tokenisasi, *stopword removal*, dan *stemming*. Setelah itu, dilakukan transformasi teks ke dalam bentuk numerik menggunakan metode TF-IDF (*Term Frequency–Inverse Document Frequency*), yang menghasilkan vektor representasi dari setiap judul skripsi. Vektor inilah yang selanjutnya digunakan dalam perhitungan jarak kemiripan menggunakan Mahalanobis Distance. Hasil perhitungan ini kemudian diuji dan dianalisis untuk menghasilkan simpulan terkait tingkat kemiripan antar judul, yang menjadi dasar dalam mendukung proses evaluasi keaslian karya ilmiah mahasiswa.



Gambar 1. Tahapan Penelitian

a) Skripsi

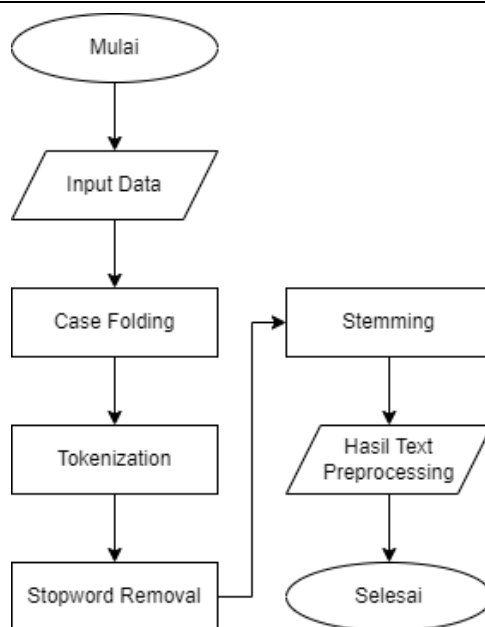
Skripsi atau tesis adalah suatu karya ilmiah yang mengemukakan pendapat penulis berdasarkan pendapat orang lain. Pendapat yang dikemukakan tidak boleh sembarangan, karena harus didukung oleh data dan fakta yang empiris dan objektif, baik berdasarkan penelitian langsung maupun penelitian tidak langsung. Biasanya ditulisnya sebuah skripsi atau karya ilmiah untuk memenuhi persyaratan guna memperoleh gelar sarjana dan penyusunannya dipandu oleh seorang dosen yang ditunjuk oleh perguruan tinggi. Dengan demikian, data atau fakta dapat didasarkan pada pengalaman empiris, hasil penelitian lapangan ataupun diambil dari literatur [1], [3], [7], [8].

Berikut ini kode dan judul skripsi hasil proses *case folding*.

- D0 : deteksi plagiarisme tingkat kemiripan judul skripsi dengan algoritma *winnowing*
- D1 : perbandingan algoritma *winnowing* dengan *rabin karp* untuk mendeteksi plagiarisme pada kemiripan teks judul skripsi
- D2 : deteksi kemiripan isi dokumen teks menggunakan algoritma *levenshtein distance*
- D3 : deteksi plagiarisme pada dokumen jurnal menggunakan metode *cosine similarity*
- D4 : deteksi tingkat kemiripan judul menggunakan algoritma *oliver*

b) Text Preprocessing

Text Preprocessing sebagai langkah awal yang digunakan untuk menyiapkan *text* sebelum diolah atau digunakan (Gambar 2) untuk tujuan pengujian dan pelatihan. *Text preprocessing* sangat penting dilakukan untuk mengolah data terutama dalam bentuk *text* [9]. *Text preprocessing* bertujuan untuk mengurangi *noise* yang terdapat pada data agar kinerja *classifier* meningkat dan proses klasifikasi juga menjadi lebih cepat [10].



Gambar 2. Tahapan Pre Processing

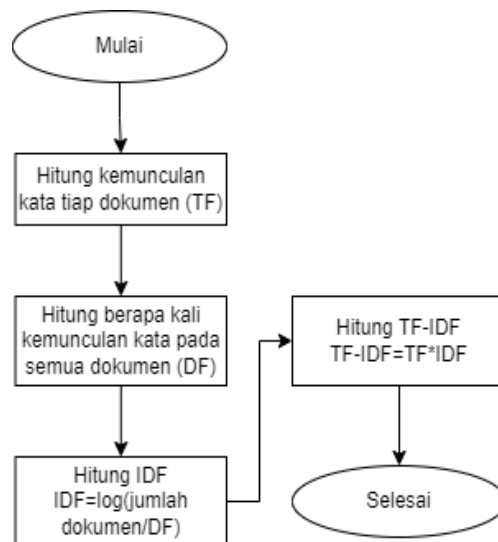
Tabel 1 menampilkan token-token yang diperoleh dari 4 judul skripsi sebelum dilakukan penghilangan *stopword*.

Tabel 1. Hasil Proses Tokenisasi

No	Token					Jumlah
D0	deteksi skripsi	plagiarisme dengan	tingkat algoritma	kemiripan <i>winnowing</i>	judul	9
D1	perbandingan karp kemiripan	algoritma untuk teks	<i>winnowing</i> mendeteksi judul	dengan plagiarisme skripsi	rabin pada	15
D2	deteksi menggunakan	kemiripan algoritma	isi <i>levenshtein</i>	dokumen <i>distance</i>	teks	9
D3	deteksi menggunakan	plagiarisme metode	pada <i>cosine</i>	dokumen <i>similarity</i>	jurnal	9
D4	deteksi algoritma	tingkat <i>oliver</i>	kemiripan	judul	menggunakan	7

c) TF-IDF

Data yang sebelumnya telah melalui tahapan *preprocessing* harus berbentuk numerik, untuk mengubah data *text* menjadi bentuk numerik dapat menggunakan metode *TF-IDF* [11], [12]. *Term Frequency Inverse Document Frequency (TF-IDF)* pada Gambar 3 adalah salah satu metode yang biasanya digunakan dalam pembobotan kata, yang dimana metode *TF-IDF* dapat menentukan nilai frekuensi sebuah kata, tujuan dari perhitungan ini yaitu untuk mengetahui seberapa relevan sebuah kata pada dokumen [8]. Metode *Term Frequency-Inverse Document Frequency (TF-IDF)* ini terkenal handal karena efisien, mudah digunakan, dan menghasilkan hasil yang akurat. Metode ini digunakan untuk menilai seberapa penting sebuah kata dalam sebuah dokumen atau sekelompok kata. Dalam dokumen tunggal, setiap kalimat dianggap sebagai dokumen. Semakin banyak dokumen yang mengandung kata tersebut, menunjukkan bahwa kata tersebut semakin umum digunakan dalam dokumen tersebut [12], [13].



Gambar 3. Tahapan TF IDF

Sebelum perhitungan TF, DF, dan IDF diperoleh kata dasar (*term*) dengan mengikuti proses *stemming* sebelumnya (Tabel 2).

Tabel 2. Kumpulan Kata Dasar (Term) Hasil Stemming

No	Term	No	Term	No	Term
1	algoritma	8	judul	15	plagiarism
2	banding	9	jurnal	16	rabin
3	cosine	10	karp	17	similarity
4	deteksi	11	levenshtein	18	skripsi
5	distance	12	metode	19	teks
6	dokumen	13	mirip	20	tingkat
7	isi	14	oliver	21	winnowing

Jumlah token pada setiap dokumen yaitu D0 = 9, D1 = 15, D2 = 9, D3 = 9, dan D4 = 7.

Perhitungan TF (*Term Frequency*) yaitu seberapa sering sebuah kata muncul dalam sebuah dokumen.

$$TF(t, d) = \frac{\text{Jumlah kemunculan term } t \text{ di dokumen } d}{\text{Total jumlah kata di dokumen } d} \quad (1)$$

Term = “algoritma” untuk masing-masing dokumen mendapatkan nilai TF yaitu:

$$TF(\text{“algoritma”}, D0) = \frac{1}{9} = 0,111$$

$$TF(\text{“algoritma”}, D1) = \frac{1}{15} = 0,067$$

$$TF(\text{“algoritma”}, D2) = \frac{1}{9} = 0,111$$

$$TF(\text{“algoritma”}, D3) = \frac{0}{9} = 0$$

$$TF(\text{“algoritma”}, D4) = \frac{1}{7} = 0,143$$

Term ini muncul pada 4 dokumen sehingga nilai IDF yaitu:

$$IDF(\text{“algoritma”}) = \log_{10}\left(\frac{5}{4}\right) = \log_{10}(1,25) = 0,0969$$

Sehingga nilai TF-IDF yaitu :

$$TF - IDF(\text{“algoritma”}, D0) = 0,111 * 0,0969 = 0,0107559$$

$$TF - IDF(\text{“algoritma”}, D1) = 0,067 * 0,0969 = 0,0064923$$

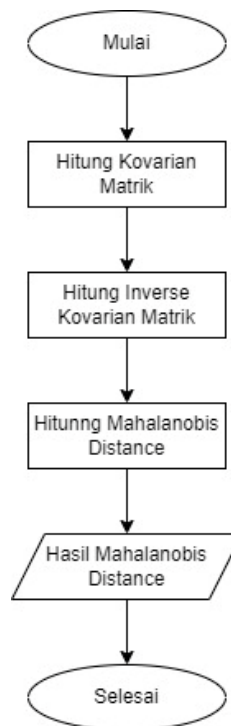
$$TF - IDF(\text{“algoritma”}, D2) = 0,111 * 0,0969 = 0,0107559$$

$TF - IDF("algoritma"), D3) = 0 * 0,0969 = 0$
 $TF - IDF("algoritma"), D4) = 0,143 * 0,0969 = 0,0138567$
 Untuk perhitungan TF-IDF selengkapnya terlihat pada Table 3.

Tabel 3. Nilai TF-IDF untuk semua term

No	Term	TF D0	TF D1	TF D2	TF D3	TF D4	IDF	TF-IDF				
		9	15	9	9	7		D0	D1	D2	D3	D4
1	algoritma	0.111	0.067	0.111	0	0.143	0.097	0.011	0.006	0.011	0	0.014
2	banding	0	0.067	0	0	0	0.699	0	0.047	0	0	0
3	cosine	0	0	0	0.111	0	0.699	0	0	0	0.078	0
4	deteksi	0.111	0.067	0.111	0.111	0.143	0	0	0	0	0	0
...		...	...		...	...	...	...	...	...	...	...
16	rabin	0	0.067	0	0	0	0.699	0	0.047	0	0	0
17	similarity	0	0	0	0.111	0	0.699	0	0	0	0.078	0
18	skripsi	0.111	0.067	0	0	0	0.398	0.044	0.027	0	0	0
19	teks	0	0.067	0.111	0	0	0.398	0	0.027	0.044	0	0
20	tingkat	0.111	0	0	0	0.143	0.398	0.044	0	0	0	0.057
21	winnowing	0.111	0.067	0	0	0	0.398	0.044	0.027	0	0	0

d) Mahalanobis Distance



Gambar 4. Tahapan Mahalanobis Distance

Gambar 4 memperlihatkan metode *Mahalanobis Distance* merupakan salah satu metode perhitungan jarak yang dapat digunakan untuk menilai seberapa mirip objek berdasarkan pada hubungan antar variabel dengan pola yang beragam. Metode ini memungkinkan pengukuran jarak data terhadap mean, sehingga memperlihatkan dengan jelas penyebaran data dengan pola terhadap nilai rata-rata [4]. Metode *Mahalanobis* memiliki keunggulan, salah satunya adalah kemampuannya untuk membandingkan kedua objek meskipun memiliki jumlah data yang tidak sama

berdasarkan korelasi variabel dengan pola yang berbeda.

Berikut adalah rumus dari *Mahalanobis Distance* [14]. Berikut adalah rumus dari *Mahalanobis Distance* :

1. Hitung *varians* dan *covariance matrix*, perhitungan *varians* dan *covariance matrix* dapat dihitung dengan algoritma berikut:

$$Var = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (2)$$

Keterangan:

x_i = Titik data

$\bar{x}$ = Rata-rata semua data

n = Banyaknya data

$$S = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})(T_i - \bar{T}) \quad (3)$$

Keterangan:

S = *Covariance matrix* dari sampel

X_i = Nilai ke i dari variabel X

T_i = Nilai ke i dari variabel T

$\bar{T}$ = *Mean* variabel T

$\bar{X}$ = *Mean* variabel X

2. Menghitung nilai inverse kovarian, inverse kovarian dapat dihitung menggunakan rumus berikut:

$$\Sigma^{-1} = (\Sigma)^{-1} \quad (4)$$

Ket:

$(\Sigma)^{-1}$ = Matrik inverse Σ

3. Menghitung jarak menggunakan rumus *Mahalanobis* berikut:

$$d_{i=1}^n = \sqrt{(\bar{X} - \bar{T})^T S^{-1} (\bar{X} - \bar{T})} \quad (5)$$

Ket:

$d_{i=1}^n$ = Jarak data

S = *Covariance matrix* dari sampel

S^{-1} = *Invers* Matriks

$\bar{T}$ = *Vector* dari sampel

$\bar{X}$ = *Data vector* uji

e) *Confusion Matrix*

Confusion matrix adalah sebuah tabel yang dapat menunjukkan jumlah data uji yang diklasifikasikan dengan benar dan jumlah data uji yang diklasifikasikan dengan salah. Dalam dunia klasifikasi, *confusion matrix* hadir sebagai alat penting untuk evaluasi performa. Dengan memanfaatkan data dan metrik yang tertanam dalam strukturnya, *Confusion matrix* memberikan gambaran tentang efektivitas proses klasifikasi. Matriks ini dengan cermat membandingkan kelas aktual dengan kelas yang diprediksi oleh matriks. Hasil yang diperoleh, yaitu sekumpulan lengkap nilai *True Positive (TP)*, *True Negative (TN)*, *False Positive (FP)*, dan *False Negative (FN)*, memberikan gambaran yang jelas tentang kekuatan dan area yang perlu ditingkatkan dari sistem [11]. Berikut rumus *confusion matrix* untuk menghitung *precision*, *recall*, *F1-score*, dan Akurasi.

$$precision = \frac{TP}{TP+FP} \quad (6)$$

$$recall = \frac{TP}{TP+FN} \quad (7)$$

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision+Recall} \quad (8)$$

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \quad (9)$$

3. HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil implementasi metode Mahalanobis Distance dalam mengidentifikasi tingkat kemiripan judul skripsi, serta pembahasan terhadap temuan-temuan yang diperoleh dari proses tersebut. Analisis dilakukan berdasarkan data judul skripsi yang telah melalui tahapan text preprocessing dan representasi vektor menggunakan TF-IDF. Selanjutnya, dilakukan perhitungan jarak kemiripan antar judul untuk mengukur sejauh mana kesamaan konten dapat dikenali oleh sistem. Hasil yang diperoleh dibandingkan dengan penilaian manual serta dianalisis untuk menilai efektivitas metode yang digunakan dalam mendeteksi kemiripan judul secara akurat dan objektif. Pembahasan juga mencakup interpretasi hasil, identifikasi kelebihan dan keterbatasan metode, serta potensi pengembangan lebih lanjut.

A. Data Penelitian

Data penelitian yang digunakan (Gambar 5) adalah judul skripsi Teknik Informatika UIGM dari tahun 2020 – 2023. Total jumlah sample data yang digunakan sebanyak 112 judul skripsi.

id	judul	jumlah	tempat	penerbit	Prodi	kode panggil
0 25204	IMPLEMENTASI CHATBOT DENGAN METODE ARTIFICIAL INTELLIGENCE MARKUP LANGUAGE (AIML) PADA PERCETAKAN MUG CANTIK	1	Palembang	Fakultas Ilmu Komputer	Teknik Informatika	E-SKR/2020/1
1 23962	OPTIMASI PENJADWALAN PERKULIAHAN FAKULTAS ILMU KOMPUTER DI UNIVERSITAS INDO GLOBAL MANDIRI PALEMBANG DENGAN METODE HARMONY SEARCH	1	Palembang	Fakultas Ilmu Komputer	Teknik Informatika	E-SKR/2020/10
2 23963	Pemanfaatan Algoritma Apriori untuk Menganalisis Kebutuhan Suku Cadang di PT Indofood CBP Sukses Makmur Tbk	1	Palembang	Fakultas Ilmu Komputer	Teknik Informatika	E-SKR/2020/11
3 23964	IMPLEMENTASI METODE HYBRID (COLLABORATIVE BASED FILTERING DAN CONTENT BASED FILTERING) UNTUK SISTEM REKOMENDASI PADA E-COMMERCE PEMESANAN HIBURAN MASYARAKAT	1	Palembang	Fakultas Ilmu Komputer	Teknik Informatika	E-SKR/2020/12
4 23965	IMPLEMENTASI LOCATION BASED SERVICE (LBS) PADA APLIKASI Pencarian Toko OLEH-OLEH DI KOTA PALEMBANG BERBASIS ANDROID	1	Palembang	Fakultas Ilmu Komputer	Teknik Informatika	E-SKR/2020/13
...	...	...	...	...	...	...
107 26643	REKAYASA PERANGKAT PEMBELAJARAN IPA MENGGUNAKAN METODE SDLC PROTOTYPE BAGIS SISWA SMP N 2 PEDAMARAN TIMUR BERBASIS FUN LEARNING	1	Palembang	Fakultas Ilmu Komputer	Teknik Informatika	E-SKR/2023/6
108 26644	IMPLEMENTASI HONEYPOT SEBAGAI PENDETEKSI SERANGAN PADA VIRTUAL PRIVATE SERVER (VPS)	1	Palembang	Fakultas Ilmu Komputer	Teknik Informatika	E-SKR/2023/7
109 26645	KOMPARASI ALGORITMA MACHINE LEARNING DALAM PEMILIHAN CALON MAHASISWA PENERIMA KIP KULIAH	1	Palembang	Fakultas Ilmu Komputer	Teknik Informatika	E-SKR/2023/8
110 26646	PENYISIPAN PESAN PADA CITRA DIGITAL MENGGUNAKAN METODE SPREAD SPECTRUM	1	Palembang	Fakultas Ilmu Komputer	Teknik Informatika	E-SKR/2023/9
111 26647	APLIKASI UJIAN ONLINE BERBASIS ANDROID VIEW UNTUK SISWA SEKOLAH MENENGAH ATAS MENGGUNAKAN METODE RATIONAL UNIFIED PROCESS (RUP)	1	Palembang	Fakultas Ilmu Komputer	Teknik Informatika	E-SKR/2023/10

Gambar 5. Data Udul Kripsi Yang Pernah Digunakan

B. Tahap Text Preprocessing

a) Case Folding dan Tokenizing

Pada tahap ini (Gambar 6a), semua huruf pada judul diubah menjadi huruf kecil. Dilanjutkan *tokenizing* judul dengan memecah *string* menjadi terpisah (Gambar 6b).

Case Folding Result	Tokenizing Result :
0 implementasi chatbot dengan metode artificial ...	0 [implementasi, chatbot, dengan, metode, artifi...
1 optimasi penjadwalan perkuliahan fakultas ilmu...	1 [optimasi, penjadwalan, perkuliahan, fakultas,...
2 pemanfaatan algoritma apriori untuk menganalis...	2 [pemanfaatan, algoritma, apriori, untuk, menga...
3 implementasi metode hybrid (collaborative bas...	3 [implementasi, metode, hybrid, collaborative, ...
4 implementasi location based service (lbs) pada...	4 [implementasi, location, based, service, lbs, ...
...	...
107 rekayasa perangkat pembelajaran ipa mengguna...	107 [rekayasa, perangkat, pembelajaran, ipa, menggg...
108 implementasi honeypot sebagai pendeteksi seran...	108 [implementasi, honeypot, sebagai, pendeteksi, ...
109 komparasi algoritma machine learning dalam pem...	109 [komparasi, algoritma, machine, learning, dala...
110 penyisipan pesan pada citra digital mengguna...	110 [penyisipan, pesan, pada, citra, digital, meng...
111 aplikasi ujian online berbasis android view un...	111 [aplikasi, ujian, online, berbasis, android, v...
Name: judul, Length: 112, dtype: object	Name: Judul_tokens, Length: 112, dtype: object

(a)

(b)

Gambar 6. Hasil Proses Case Folding (a) dan Tokenisasi Judul Skripsi (b)

b) Stopword dan Stemming

Pada tahap ini dilakukan *stopword* pada judul untuk menghilangkan kata-kata yang tidak penting yang

diperlihatkan pada Gambar 7a. Kemudian dilakukan *stemming* untuk mengubah kata menjadi bentuk dasarnya (Gambar 7b).

<pre> Judul Tokens 0 [implementasi, chatbot, dengan, metode, artifi... 1 [optimasi, penjadwalan, perkuliahan, fakultas,... 2 [pemanfaatan, algoritma, apriori, untuk, menga... 3 [implementasi, metode, hybrid, collaborative, ... 4 [implementasi, location, based, service, lbs, 107 [rekayasa, perangkat, pembelajaran, ipa, mengg... 108 [implementasi, honeypot, sebagai, pendeteksi, ... 109 [komparasi, algoritma, machine, learning, dala... 110 [penyisipan, pesan, pada, citra, digital, meng... 111 [aplikasi, ujian, online, berbasis, android, v... Name: Judul_tokens, Length: 112, dtype: object Hasil Stopword 0 [implementasi, chatbot, metode, artificial, in... 1 [optimasi, penjadwalan, perkuliahan, fakultas,... 2 [pemanfaatan, algoritma, apriori, menganalisis... 3 [implementasi, metode, hybrid, collaborative, ... 4 [implementasi, location, based, service, lbs, 107 [rekayasa, perangkat, pembelajaran, ipa, metod... 108 [implementasi, honeypot, pendeteksi, serangan,... 109 [komparasi, algoritma, machine, learning, pemi... 110 [penyisipan, pesan, citra, digital, metode, sp... 111 [aplikasi, ujian, online, berbasis, android, v... Name: judul_tokens_WSW, Length: 112, dtype: object </pre>	<pre> 634 ----- implementasi : implementasi chatbot : chatbot metode : metode artificial : artificial intelligence : intelligence markup : markup language : language aiml : aiml percetakan : cetak mug : mug cantik : cantik optimasi : optimasi penjadwalan : jadwal perkuliahan : kuliah fakultas : fakultas ilmu : ilmu komputer : komputer universitas : universitas indo : indo global : global mandiri : mandiri palembang : palembang harmony : harmony ... 109 [komparasi, algoritma, machine, learning, pili... 110 [sisip, pesan, citra, digital, metode, spread,... 111 [aplikasi, uji, online, bas, android, view, si... Name: judul_tokens_stemmed, Length: 112, dtype: object </pre>
(a)	(b)

Gambar 7. Hasil Stopword Setiap Judul Skripsi (a), dan Stopword Setiap Judul Skripsi (b)

C. Tahap Pembobotan

a) Hitung Term Frequency

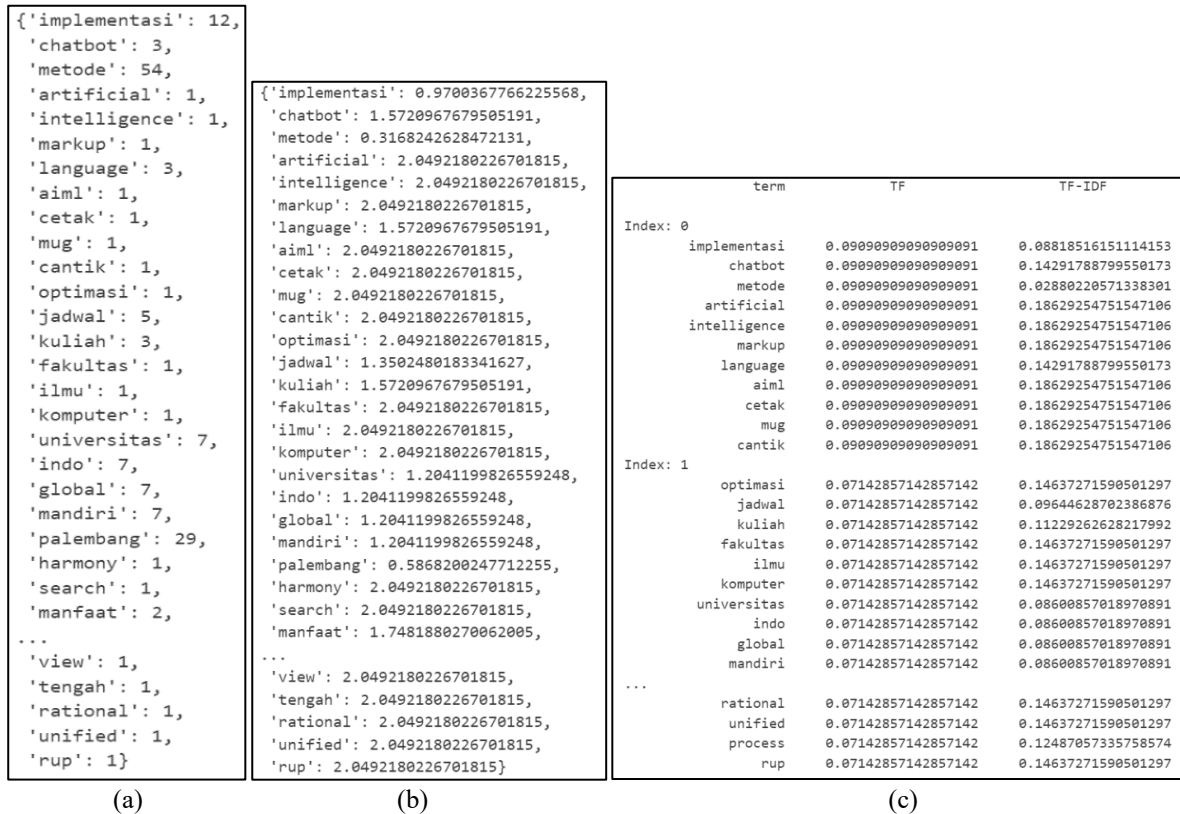
Pada tahap ini menghitung seberapa sering suatu kata muncul dalam sebuah judul (Gambar 8).

term	TF
Index: 0	
implementasi	0.09090909090909091
chatbot	0.09090909090909091
metode	0.09090909090909091
artificial	0.09090909090909091
intelligence	0.09090909090909091
markup	0.09090909090909091
language	0.09090909090909091
aiml	0.09090909090909091
cetak	0.09090909090909091
mug	0.09090909090909091
cantik	0.09090909090909091
Index: 1	
optimasi	0.07142857142857142
jadwal	0.07142857142857142
kuliah	0.07142857142857142
fakultas	0.07142857142857142
ilmu	0.07142857142857142
komputer	0.07142857142857142
universitas	0.07142857142857142
indo	0.07142857142857142
global	0.07142857142857142
...	
unified	0.07142857142857142
process	0.07142857142857142
rup	0.07142857142857142

Gambar 8. Hasil Stopword Setiap Judul Skripsi (a), dan Stopword Setiap Judul Skripsi (b)

b) Hasil Perhitungan *Document Frequency*, *Inverse Document Frequency*, dan *TF-IDF*

Pada tahap ini diperoleh berapa jumlah kemunculan sebuah kata pada seluruh judul. Menghitung *IDF* untuk mengukur seberapa penting suatu kata, dan terakhir mendapatkan nilai *TF-IDF* (Gambar 9).

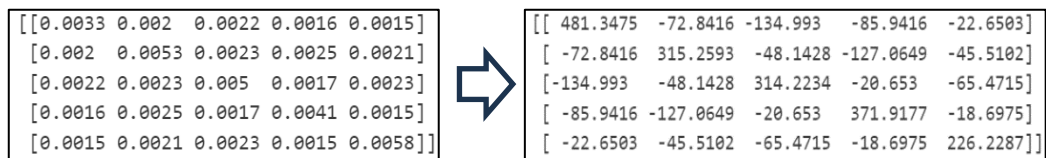


Gambar 9. Hasil *DF* (a), *IDF* (b), dan *TF-IDF* (b)

D. Tahap *Mahalanobis Distance*

a) Nilai Kovarian Matrix

Pada tahap ini dihitung kovarian matrik dari nilai vektor *TF-IDF* semua judul dan selanjutnya menghitung *inverse* dari kovarian matriknya (Gambar 10).



Gambar 10. Matriks Kovarian dan Invers-nya

b) Nilai *Mahalanobis Distance*

Selanjutnya hitung *Mahalanobis Distance*, judul skripsi pada index ke 108 akan dihitung jarak kemiripannya ke seluruh judul.

Index		Judul	Distance
108	108	sistem pendukung keputusan penerimaan pegawai pada sma bakti ibu Palembang menggunakan metode multi attribute utility theory maut	0.000000
106	106	sistem pendukung keputusan penerimaan bantuan untuk warga miskin dengan metode algoritma expectation maximization pada dinas sosial kota Palembang	0.107244
105	105	sistem pendukung keputusan penerima bantuan langsung tunai blt dana desa menggunakan best worst method bwm	0.236888
104	104	sistem pendukung keputusan penentuan dosen pembimbing skripsi menggunakan metode analytical heararchy process studi kasus universitas indo global mandiri Palembang	0.508207
107	107	sistem pendukung keputusan penerimaan beasiswa di sma bakti ibu Palembang metode ranking based on optimal points rbop	0.549310
...	...	...	...
10	10	aplikasi penerjemahan bahasa Indonesia ke bahasa Lampung dan bahasa Inggris menggunakan metode natural language processing berbasis android	5.312174
20	20	edutainment pengenalan jenis burung menggunakan augmented reality	5.414766
62	62	penyisipan pesan pada citra digital menggunakan metode spread spectrum	5.488797
34	34	klasifikasi kinerja kurir menggunakan metode na	6.118050
80	80	rancangan game avoid virus menggunakan unity	6.399373

Gambar 11. Hasil Perhitungan Mahalanobis Distance

Dapat dilihat pada Gambar 11 semakin kecil jarak *Mahalanobis*, maka semakin mirip judul tersebut, sebaliknya semakin besar jarak *Mahalanobis*, maka semakin tidak mirip judul tersebut.

E. Hasil Pengujian

a) Pengujian *Mahalanobis*

Pengujian menggunakan 6 judul yang dipilih acak akan dihitung kemiripannya dengan seluruh judul. Pengujian ini bertujuan untuk melihat *score* yang dihasilkan *Mahalanobis*.

Judul Pertama: “Aplikasi *chatbot information service center* pada PT. Agrotans Distribusi Lestari dengan pendekatan *Natural Language Processing*”.

Index		Judul	Distance
1	3	aplikasi chatbot information service center pada pt agrotans distribusi lestari dengan pendekatan natural language processing nlp	0.000000
2	14	aplikasi peramalan penjualan alat kesehatan dengan metode single moving average dan single exponential smoothing studi kasus pt parit panjang Palembang	0.651656
3	9	aplikasi pemetaan ancaman kebakaran hutan dan lahan di wilayah kabupaten penulak abab lematang ilir menggunakan location based service lbs	0.678331
4	15	aplikasi potensi kawasan wilayah ancaman bencana banjir di kabupaten ogan komering ilir menggunakan algoritma floyd warshall	0.770233
5	33	implementasi two factor authentication fa berbasis timebased one time password totp sebagai metode keamanan ganda	0.829308
...	...	...	...
108	96	simulasi pendeteksi gawang menggunakan metode harris corner detection	5.018095
109	21	edutainment pengenalan jenis burung menggunakan augmented reality	5.451228
110	64	penyisipan pesan pada citra digital menggunakan metode spread spectrum	5.709297
111	35	klasifikasi kinerja kurir menggunakan metode na	5.795785
112	82	rancangan game avoid virus menggunakan unity	5.894749

Gambar 12. Pengujian Judul Pertama untuk 5 Judul terdekat

Judul Kedua: “Klasifikasi penerima bantuan pada warga yang terdampak covid menggunakan *K-Nearest Neighbor*”.

Index	Judul	Distance
1 39	klasifikasi penerima bantuan pada warga yang terdampak covid menggunakan knearest neighbor knn	0.000000
2 19	deteksi pola wajah peserta tryout menggunakan metode eigenface studi kasus sdn banyuasin	1.192996
3 88	rekayasa aplikasi pencarian daerah genangan air dan rawan demam berdarah menggunakan algoritma haversine	1.266782
4 89	rekayasa aplikasi pengenalan jenis kopi lokal sumatera selatan berbasis android	1.327968
5 46	pemetaan daerah rawan kriminalitas menggunakan kmeans clustering polrestabes Palembang	1.329355
...	...	...
108 34	klasifikasi karakteristik kepribadian siswa menggunakan metode naive bayes	4.772472
109 64	penyisipan pesan pada citra digital menggunakan metode spread spectrum	4.949948
110 62	pengenalan wajah menggunakan algoritma viola jones untuk identifikasi wajah mahasiswa tampak depan	5.017208
111 10	aplikasi penerjemahan bahasa indonesia ke bahasa lampung dan bahasa inggris menggunakan metode natural language processing berbasis android	5.339276
112 35	klasifikasi kinerja kurir menggunakan metode na	5.522914

Gambar 13. Pengujian Judul Kedua untuk 5 Judul terdekat

Judul Ketiga: “Penyisipan pesan pada citra digital menggunakan metode *spread spectrum*”

Index	Judul	Distance
1 64	penyisipan pesan pada citra digital menggunakan metode spread spectrum	0.000000
2 17	clustering penilaian kinerja dosen dengan metode agglomerative hierarchical clustering studi kasus universitas indo global mandiri	3.168927
3 18	clustering status gizi balita menggunakan algoritma kmeans studi kasus puskesmas ogan ilir	3.257466
4 34	klasifikasi karakteristik kepribadian siswa menggunakan metode naive bayes	3.352917
5 65	peramalan jumlah kendaraan pada tol palindra Palembang indralaya menggunakan metode automatic clustering fuzzy logical relationship	3.544941
...	...	...
108 96	simulasi pendeteksi gawang menggunakan metode harris corner detection	6.358213
109 12	aplikasi pengenalan anatomi sistem pencernaan pada tubuh manusia menggunakan teknologi augmented reality	6.488065
110 35	klasifikasi kinerja kurir menggunakan metode na	7.099728
111 10	aplikasi penerjemahan bahasa indonesia ke bahasa lampung dan bahasa inggris menggunakan metode natural language processing berbasis android	7.424244
112 5	aplikasi game edukasi pengenalan warna pada anak usia dini melalui warna kendaraan	8.770385

Gambar 14. Pengujian Judul Ketiga untuk 5 Judul terdekat

Judul 4: “Prediksi calon mahasiswa penerima kip menggunakan algoritma *decision tree* pada Universitas Indo Global Mandiri”.

Index	Judul	Distance
1 70	prediksi calon mahasiswa penerima kip menggunakan algoritma decision tree pada universitas indo global mandiri	0.000000
2 29	implementasi location based service lbs pada aplikasi pencarian toko oleholeh di kota Palembang berbasis android	0.609173
3 75	prediksi permintaan mata kuliah pada semester pendek dengan menggunakan metode fuzzy time series pada program studi teknik informatika	0.740152
4 74	prediksi penjualan sepeda motor pada yamaha thamin brothers paju menggunakan algoritma neural network pada data time series	0.858294
5 107	sistem pendukung keputusan penerima bantuan langsung tunai btltdana desa menggunakan best worst method bwm	0.940178
...	...	...
108 21	edutainment pengenalan jenis burung menggunakan augmented reality	5.003719
109 64	penyisipan pesan pada citra digital menggunakan metode spread spectrum	5.018461
110 10	aplikasi penerjemahan bahasa indonesia ke bahasa lampung dan bahasa inggris menggunakan metode natural language processing berbasis android	5.283390
111 35	klasifikasi kinerja kurir menggunakan metode na	5.328514
112 82	rancangan game avoid virus menggunakan unity	5.946728

Gambar 15. Pengujian Judul Keempat untuk 5 Judul terdekat

Judul 5: “Rekayasa *virtual tour* museum negeri Sumatera Selatan menggunakan panorama berbasis web”.

Index	Judul	Distance
1 95	rekayasa virtual tour museum negeri sumatera selatan menggunakan panorama berbasis web	0.000000
2 33	implementasi two factor authentication fa berbasis timebased one time password totp sebagai metode keamanan ganda	0.727289
3 76	prediksi tingkat prevalensi gizi buruk di puskesmas babat menggunakan metode naive bayes	0.761845
4 53	penerapan internet of things iot pada monitoring suhu menggunakan logika fuzzy berbasis android	0.891767
5 0	analisis sentimen semantik top marketplace di indonesia berdasarkan promosi menggunakan recurrent neural network	0.979101
...	...	...
108 82	rancangan game avoid virus menggunakan unity	4.726470
109 62	pengenalan wajah menggunakan algoritma viola jones untuk identifikasi wajah mahasiswa tampak depan	4.731495
110 5	aplikasi game edukasi pengenalan warna pada anak usia dini melalui warna kendaraan	4.745646
111 96	simulasi pendeteksi gawang menggunakan metode harris corner detection	4.879253
112 35	klasifikasi kinerja kurir menggunakan metode na	5.613849

Gambar 15. Pengujian Judul Kelima untuk 5 Judul terdekat

Judul 6: “Sistem pendukung keputusan penerimaan pegawai pada sma bakti ibu palembang menggunakan metode *multi attribute utility theory* maut”.

Index	Judul	Distance
1 110	sistem pendukung keputusan penerimaan pegawai pada sma bakti ibu palembang menggunakan metode multi attribute utility theory maut	0.000000
2 107	sistem pendukung keputusan penerima bantuan langsung tunai btltdana desa menggunakan best worst method bwm	0.157721
3 109	sistem pendukung keputusan penerimaan beasiswa di sma bakti ibu palembang metode ranking based on optimal points rbop	0.242595
4 108	sistem pendukung keputusan penerimaan bantuan untuk warga miskin dengan metode algoritma expectation maximization pada dinas sosial kota palembang	0.270225
5 105	sistem pendukung keputusan pemilihan perumahan bersubsidi di kota palembang menggunakan best worst method bwm	0.424235
...	...	...
108 10	aplikasi penerjemahan bahasa indonesia kebahasa lampung dan bahasa inggris menggunakan metode natural language processing berbasis android	5.327991
109 21	edutainment pengenalan jenis burung menggunakan augmented reality	5.371458
110 64	penyisipan pesan pada citra digital menggunakan metode spread spectrum	5.667565
111 35	klasifikasi kinerja kurir menggunakan metode na	6.053879
112 82	rancangan game avoid virus menggunakan unity	6.281216

Gambar 16. Pengujian Judul Keenam untuk 5 Judul terdekat

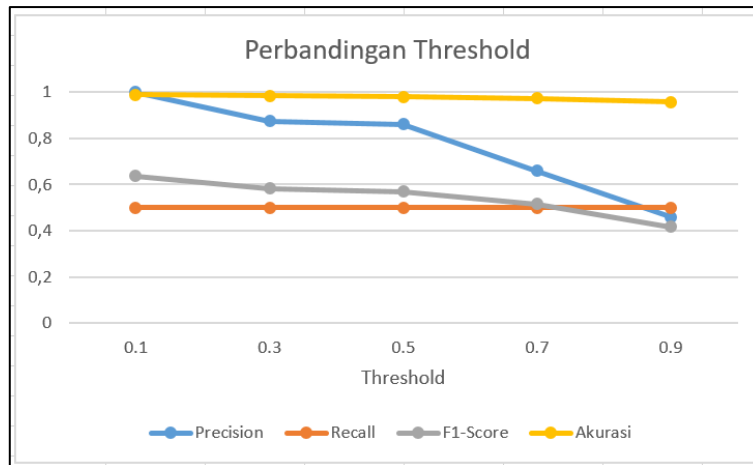
Berdasarkan pengujian Mahalanobis dapat dilihat bahwa semakin kecil nilai *Mahalanobis* maka semakin dekat jarak kemiripannya.

b) Pengujian *Threshold*

Pengujian *threshold* dilakukan untuk menentukan nilai batas yang digunakan untuk menentukan mirip atau tidaknya suatu judul. Nilai *threshold* yang diuji berjumlah 5 nilai, yaitu 0.1, 0.3, 0.5, 0.7, dan 0.9.

Tabel 1 Rata-rata Pengujian *Threshold*

Nilai <i>Threshold</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>	Akurasi
0.1	1	0.5	0.638	0.988
0.3	0.875	0.5	0.583	0.983
0.5	0.861	0.5	0.569	0.980
0.7	0.659	0.5	0.516	0.973
0.9	0.461	0.5	0.417	0.958



Gambar 17. Grafik Perbandingan Threshold Evaluasi

Berdasarkan rata-rata pengujian *threshold* dapat dilihat bahwa rata-rata nilai *precision* terbesar didapat dari *threshold* 0.1, yaitu sebesar 1,0. Sedangkan untuk nilai *precision* terkecil didapat dari *threshold* 0.9, yaitu sebesar 0,461. Pada Nilai *recall* yang didapat dari *threshold* 0.1, 0.3, 0.5, 0.7, dan 0.9 menunjukkan tidak ada perubahan nilai, semuanya memiliki rata-rata nilai *recall* yang sama yaitu sebesar 0,5. Nilai *F1-score* terbesar didapat dari *threshold* 0.1 yaitu sebesar 0,638. Sedangkan nilai *F1-score* terkecil didapat dari *threshold* 0.9 yaitu sebesar 0,417. Nilai akurasi terbesar didapat dari *threshold* 0.1 yaitu sebesar 0,988. Sedangkan nilai akurasi terkecil didapat dari *threshold* 0.9 yaitu sebesar 0,958.

Dilihat pada hasil pengujian *threshold*, hasil akurasi dari ke lima *threshold* memiliki nilai yang bagus. Semua *threshold* mendapat nilai akurasi diatas 0,95 semua. Namun jika dilihat dari nilai *precision*, *recall*, dan *F1-score* dari nilai *threshold* 0.9 yang dimana memiliki nilai *precision*, *recall*, dan *F1-score* terkecil tapi nilai akurasinya tetap tinggi. Dapat dilihat bahwa ada bias pada data atau data tidak seimbang. Hal tersebut dapat terjadi dikarenakan ada potensi bias, yang dimana pada data pasangan judul yang mirip jauh lebih sedikit jika dibandingkan dengan pasangan judul yang tidak mirip. Artinya model dapat mencapai akurasi tinggi dengan mengklasifikasikan sebagian besar judul sebagai tidak mirip karena sebagian besar judulnya memang tidak memiliki kemiripan.

Berdasarkan grafik perbandingan antar *threshold* menunjukkan bahwa semakin besar *threshold*, semakin kecil nilai *precision*, *F1-score*, dan akurasi. Sebaliknya semakin kecil nilai *threshold*, semakin besar nilai *precision*, *F1-score*, dan akurasi. Dari ke lima nilai *threshold* yang telah di uji, dipilih *threshold* 0.1. Nilai tersebut dipilih karena nilai *precision*, *F1-score*, dan akurasi memiliki nilai tertinggi dibandingkan nilai *threshold* yang lain. Dengan kata lain jika jarak kemiripan kurang dari 0.1 maka judul dapat dikatakan mirip

4. KESIMPULAN

Hasil evaluasi menunjukkan bahwa metode Mahalanobis Distance belum memberikan kinerja optimal dalam identifikasi kemiripan judul skripsi. Meskipun *precision* mencapai nilai tertinggi 1,0, nilai *recall* yang rendah (0,5) dan *F1-score* yang terbatas (maksimal 0,638) menandakan efektivitas yang kurang dalam mendeteksi kesamaan secara menyeluruh. Akurasi yang tinggi (0,958–0,988) tidak cukup merepresentasikan keberhasilan metode dalam konteks klasifikasi sensitif. Selain itu, grafik perbandingan menunjukkan bahwa peningkatan nilai *threshold* cenderung menurunkan *precision*, *F1-score*, dan akurasi, sedangkan penurunan *threshold* justru meningkatkan ketiganya. Oleh karena itu, diperlukan pendekatan alternatif atau metode gabungan untuk meningkatkan performa sistem secara keseluruhan.

DAFTAR PUSTAKA

- [1] M. Wangge, "Penerapan Metode Principal Component Analysis (PCA) Terhadap Faktor-faktor yang Mempengaruhi Lamanya Penyelesaian Skripsi Mahasiswa Program Studi Pendidikan Matematika FKIP UNDANA," vol. 05, no. 02, pp. 974–988, 2021.
- [2] W. Dafit Prayoga, M. Bakri, and Y. Rahmanto, "Aplikasi Perpustakaan Berbasis OPAC (Online Public Access Catalog) Di Smk N 1 Talangpadang," *Jurnal Informatika dan Rekayasa Perangkat Lunak (JATIKA)*, vol. 1, no. 2, pp. 183–191, 2020, [Online]. Available: <http://jim.teknokrat.ac.id/index.php/informatika>
- [3] H. Eko Ujiantoro Adi Basuki, K. Kusri, and A. Nasiri, "Studi Dan Analisis Algoritma Kemiripan Cosine Similarity Untuk Mengidentifikasi Kesamaan Judul Skripsi," in *Prosiding Seminar Ilmiah Sistem Informasi Dan Teknologi Informasi*, 2023, pp. 1–9.
- [4] T. Kantacarini, D. Erny Herwindiaty, and J. Hendryli, "Kinerja Metode Mahalanobis Distance Yang Dibentuk Dari Dua Ukuran Pusat Dan Dua Dispersi Multivariat (Untuk Ukuran Similaritas Klasifikasi Image)," *Jurnal Ilmu Komputer dan Sistem Informasi*, vol. 9, no. 1, pp. 1–4, 2021.
- [5] D. Ardiyansyah and N. Oktafiani, "Perbandingan Metode Pengukuran Jarak Pada K-Nearest Neighbour Dalam Klasifikasi Data Teks Kardiovaskular," *Journal of Information Systems Management and Digital Business (JISMDB)*, vol. 1, no. 2, pp. 116–122, 2024.
- [6] J. H. Hadari Nawawi, "Identifikasi Sidik Jari Menggunakan Metode Principal Component Analysis (Pca) Dan Mahalanobis Distance Pada Pemilihan Presiden Mahasiswa," *Jurnal Komputer dan Aplikasi*, vol. 07, no. 02, pp. 40–48, 2019.
- [7] B. Harira Irawan, M. H. Sahat Simarangkir, and S. MIC Cikarang, "Deteksi Kemiripan Judul Skripsi Menggunakan Algoritma Levenshtein Distance Pada Kampus Stmik Mic Cikarang," *Jurnal Ilmiah Educat*, vol. 7, no. 2, pp. 143–149, 2021.
- [8] M. Wangge, "Penerapan Metode Principal Component Analysis (PCA) Terhadap Faktor-faktor yang Mempengaruhi Lamanya Penyelesaian Skripsi Mahasiswa Program Studi Pendidikan Matematika FKIP UNDANA," *Jurnal Cendekia: Jurnal Pendidikan Matematika*, vol. 05, no. 02, pp. 974–988, 2021.
- [9] L. Hermawan, M. B. Ismiati, J. Bangau, N. 60, and M. Charitas, "Pembelajaran Text Preprocessing berbasis Simulator Untuk Mata Kuliah Information Retrieval," *TRANSFORMATIKA*, vol. 17, no. 2, pp. 188–199, 2020.
- [10] M. U. Albab, Y. K. P., and M. N. Fawaiq, "Optimization of the Stemming Technique on Text Preprocessing President 3 Periods Topic," *Jurnal Transformatika*, vol. 20, no. 2, pp. 1–12, Jan. 2023, doi: 10.26623/transformatika.v20i2.5374.
- [11] O. I. Gifari, M. Adha, I. Rifky Hendrawan, F. Freddy, and S. Durrand, "Analisis Sentimen Review Film Menggunakan TF-IDF dan Support Vector Machine," *JIFOTECH (Journal Of Information Technology)*, vol. 2, no. 1, 2022.
- [12] J. Andre Septian, T. Maulana Fahrudin, and A. Nugroho, "Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF-IDF dan K-Nearest Neighbor," *Journal Of Intelligent Systems And Computation*, [Online]. Available: <https://t.co/9WloaWpfD5>
- [13] J. Andre Septian, T. Maulana Fahrudin, and A. Nugroho, "Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF-IDF dan K-Nearest Neighbor," *Journal Of Intelligent Systems And Computation*, vol. 1, no. 1, pp. 43–49, 2019, [Online]. Available: <https://t.co/9WloaWpfD5>
- [14] M. N. Kadhim, D. Al-Shammary, A. M. Mahdi, and A. Ibaida, "Feature selection based on Mahalanobis distance for early Parkinson disease classification," *Computer Methods and Programs in Biomedicine Update*, vol. 7, Jan. 2025, doi: 10.1016/j.cmpbup.2025.100177.