



**Identifikasi Kemiripan Judul Skripsi Menggunakan Metode
Mahalanobis Distance**

SKRIPSI

**Diajukan Sebagai Syarat Untuk Menyelesaikan
Pendidikan Program Studi Strata-1 Pada
Program Studi informatika**

Oleh:

**Muhammad Fajri Munawar
2019.11.0028**

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER DAN SAINS
UNIVERSITAS INDO GLOBAL MANDIRI
2024**

**Identifikasi Kemiripan Judul Skripsi Menggunakan Metode
Mahalanobis Distance**



SKRIPSI

**Diajukan Sebagai Syarat Untuk Menyelesaikan
Pendidikan Program Studi Strata-1 Pada
Program Studi informatika**

Oleh:

**Muhammad Fajri Munawar
2019.11.0028**

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER DAN SAINS
UNIVERSITAS INDO GLOBAL MANDIRI
2024**

LEMBAR PENGESAHAN SKRIPSI

LEMBAR PENGESAHAN SKRIPSI

Identifikasi Kemiripan Judul Skripsi Menggunakan Metode Mahalanobis Distance

Oleh

Muhammad Fajri Munawar

NPM : 2019.11.0028

Palembang, 19 Juli 2024

Pembimbing I



Rudi Heriansyah, S.T., M.Eng., Ph.D
NIK : 2022.01.0315

Pembimbing II



Muhammad Haviz Irfani, S.Si., M.T.I
NIK:2021.03.0291

Mengetahui,

Dekan Fakultas Ilmu Komputer dan Sains

FAKULTAS ILMU KOMPUTER DAN SAINS

UIK



Rudi Heriansyah, S.T., M.Eng., Ph.D
NIK: 2022.01.0315

LEMBAR PERSETUJUAN DEWAN PENGUJI

LEMBAR PERSETUJUAN DEWAN PENGUJI

Pada hari Jumat tanggal 5 Juli 2024 telah dilaksanakan ujian sidang skripsi :

Nama : Muhammad Fajri Munawar

NPM : 2019.11.0028

Judul : Identifikasi Kemiripan Judul Skripsi Menggunakan Metode Mahalanobis Distance

Oleh Prodi Teknik Informatika Fakultas Ilmu Komputer dan Sains Universitas Indo Global Mandiri Palembang

Palembang, 7 Juli 2024

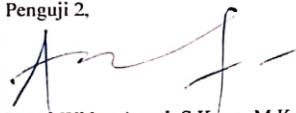
Penguji 1,



Dr. Shinta Puspasari, S.Si., M.Kom

NIK: 2015.01.0132

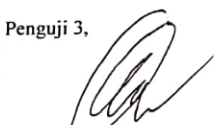
Penguji 2,



Lastri Widya Astuti, S.Kom., M.Kom

NIK: 2003.01.0063


Penguji 3,



Muhammad Haviz Irfani, S.Si., M.T.I

NIK: 2021.03.0291

Menyetujui,
Ka. Prodi Teknik Informatika



Zaid Romegar Mar, S.T., M.Cs
NIK: 2021.01.0307

SURAT KETERANGAN REVISI SKRIPSI



SURAT KETERANGAN REVISI SKRIPSI
PROGRAM STUDI TEKNIK INFORMATIKA (S1)
FASILKOM DAN SAINS UNIVERSITAS INDO GLOBAL MANDIRI

Kami yang bertanda tangan dibawah ini, menerangkan bahwa :

Nama : Muhammad Fajri Munawar
NPM : 2019.11.0028
Judul : Identifikasi Kemiripan Judul Skripsi Menggunakan Metode Mahalanobis Distance

Mahasiswa yang namanya tercantum diatas, telah selesai merevisi penulisan SKRIPSI

Palembang, 7 Juli 2024

Penguji 1,

Dr. Shima Puspasari, S.Si., M.Kom
NIK: 2015.01.0132

Penguji 2,

Lastri Widya Astuti, S.Kom., M.Kom
NIK: 2003.01.0063

Penguji 3,

Muhammad Haviz Infani, S.Si., M.T.I
NIK: 2021.03.0291

Menyetujui,
Ka. Prodi Teknik Informatika

Zaid Romegar Mair, S.T., M.Cs
NIK: 2021.01.0307

Identifikasi Kemiripan Judul Skripsi Menggunakan Metode Mahalanobis Distance

ABSTRAK

Penelitian ini bertujuan untuk mengidentifikasi kemiripan judul skripsi menggunakan metode *Mahalanobis Distance*. Metode ini telah terbukti efektif dalam mengukur kesamaan antara dua vektor ataupun distribusi data, dan metode ini dapat diterapkan dalam konteks menghitung kemiripan judul skripsi. Berdasarkan hasil pengujian Mahalanobis dalam menghitung jarak kemiripan, Metode ini mampu menghasilkan nilai kemiripan antar judul. Metode ini juga dapat diterapkan untuk mengidentifikasi kemiripan judul skripsi. Walaupun dapat diterapkan untuk identifikasi kemiripan antar judul skripsi, metode ini kurang efektif dalam melakukan identifikasi kemiripan antar judul. Dengan nilai *precision* yang mencapai 1,0, artinya metode ini tergolong andal ketika mengidentifikasi dua judul sebagai mirip. Namun untuk *recall* nilainya tergolong kecil hanya mencapai 0,5, artinya metode ini juga ada gagal dalam mengidentifikasi pasangan judul yang seharusnya mirip, namun tidak terdeteksi. Sedangkan untuk nilai *F1-score* hanya mencapai 0,638. Untuk akurasi tertinggi mencapai 0,988. Berdasarkan pengujian nilai *threshold*, nilai *threshold* 0.1 dipilih karena paling optimal dibandingkan dengan nilai *threshold* yang lain.

Kata Kunci: Skripsi, *Mahalanobis Distance*, identifikasi, *precision*, *recall*, *F1-score*, akurasi, *threshold*

Identify Similarities in Thesis Titles Using the Mahalanobis Distance Method

ABSTRACT

This research aims to identify similarities in thesis titles using the Mahalanobis Distance method. This method has been proven effective in measuring the similarity between two vectors or data distributions, and this method can be applied in the context of calculating the similarity of thesis titles. Based on the Mahalanobis test results in calculating the similarity distance, this method is able to produce similarity values between titles. This method can also be applied to identify similarities in thesis titles. Although it can be applied to identify similarities between thesis titles, this method is less effective in identifying similarities between titles. With a precision value of 1.0, this means that this method is relatively reliable when identifying two titles as similar. However, the recall value is relatively small, only reaching 0.5, meaning that this method also failed to identify pairs of titles that should be similar, but were not detected. Meanwhile, the F1-score value only reached 0.638. The highest accuracy reaches 0.988. Based on testing threshold values, a threshold value of 0.1 was chosen because it is the most optimal compared to other threshold values.

Keywords: *Thesis, Mahalanobis Distance, identification, precision, recall, F1-score, accuracy, threshold*

KATA PENGANTAR

Puji dan Syukur Penulis persembahkan kepada Tuhan Yang Maha Esa karena akhirnya skripsi ini bisa terselesaikan dengan baik tepat pada waktunya. Skripsi yang Penulis buat dengan judul **Identifikasi Kemiripan Judul Skripsi Menggunakan Metode Mahalanobis Distance** dibuat sebagai salah satu syarat untuk menyelesaikan pendidikan program studi strata-1 pada program studi informatika.

Tidak lupa Penulis mengucapkan terima kasih atas bantuan yang diberikan selama penyusunan skripsi ini kepada :

1. Dr. Marzuki Alie, SE., MM, selaku Rektor Universitas Indo Global Mandiri Palembang.
2. Rudi Heriansyah, ST., M.Eng.Ph.D, selaku Dekan Fakultas Ilmu Komputer dan pembimbing I yang telah meluangkan waktu, membimbing dan memberikan arahan kepada penulis dalam proses menyelesaikan skripsi ini.
3. Zaid Romegar Mair, S.Y., M.Cs, selaku Ketua Program Studi Teknik Informatika.
4. Shinta Puspasari, S. Si., M. Kom selaku Pembimbing Akademik saya.
5. Muhammad Haviz Irfani, S.Si.,M.T.I, selaku pembimbing skripsi II yang telah meluangkan waktu untuk membimbing penulis dan memberikan motivasi kepada penulis dalam proses menyelesaikan skripsi ini.
6. Dosen-dosen yang ada di Fakultas Ilmu Komputer Universitas IGM.
7. Teristimewa kepada kedua orang tua penulis, Bapak Yusman dan Ibu Nur Fatimah yang telah memberikan dukungan semangat serta doa sehingga penulis dapat menyelesaikan skripsi ini, dan juga untuk adik penulis yang selalu mendoakan dan memberikan dukungan kepada penulis.
8. Teman-teman penulis yang selalu membantu dan memberikan dukungan kepada penulis.

Penulis menyadari bahwa penyusunan skripsi ini masih memiliki banyak kekurangan, karenanya Penulis mengharapkan saran dan kritik yang sifatnya membangun agar dapat digunakan demi perbaikan skripsi ini nantinya. Penulis juga

berharap agar skripsi akan memberikan banyak manfaat bagi yang membacanya.

Penulis,

DAFTAR ISI

HALAMAN JUDUL

LEMBAR PENGESAHAN SKRIPSI	iii
LEMBAR PERSETUJUAN DEWAN PENGUJI	iv
SURAT KETERANGAN REVISI SKRIPSI	v
ABSTRAK	vi
<i>ABSTRACT</i>	vii
KATA PENGANTAR	viii
DAFTAR ISI	x
DAFTAR TABEL	xiii
DAFTAR GAMBAR	xiv
DAFTAR LAMPIRAN	xv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah	2
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	3
1.6 Sistematika Penulisan	3
BAB II LANDASAN TEORI	5
2.1 Skripsi	5
2.2 <i>Text Mining</i>	5
2.3 <i>Text Preprocessing</i>	5
2.3.1 Case Folding	6
2.3.2 Tokenizing	6
2.3.3 Stopword Removal	7
2.3.4 Stemming	7
2.4 <i>Term Frequency Inverse Document Frequency (TF-IDF)</i>	8
2.5 <i>Mahalanobis Distance</i>	9
2.6 <i>Flowchart</i>	11
2.7 <i>Confusion Matrix</i>	12

2.8	<i>Python</i>	14
2.9	<i>Python Library</i>	15
2.9.1	Natural Language Toolkit (NLTK).....	15
2.9.2	Sastrawi	16
2.10	Penelitian Terdahulu	17
BAB III METODE PENELITIAN.....		22
3.1	Tahapan Penelitian	22
3.2	Studi Literatur.....	22
3.3	Pengumpulan Data	22
3.4	Analisis Masalah	23
3.5	<i>Text Preprocessing</i>	23
3.5.1	<i>Input Data</i>	24
3.5.2	<i>Case Folding</i>	24
3.5.3	<i>Tokenizing</i>	25
3.5.4	<i>Stopword Removal</i>	26
3.5.5	<i>Stemming</i>	27
3.5.6	Hasil <i>Text Preprocessing</i>	27
3.6	<i>Term Frequency Inverse Document Frequency</i>	28
3.6.1	Hitung <i>Term Frequency (TF)</i>	29
3.6.2	Hitung <i>Document Frequency (DF)</i>	30
3.6.3	Hitung <i>Inverse Document Frequency (IDF)</i>	31
3.6.4	Hitung <i>Term Frequency Inverse Document Frequency (TF-IDF)</i> .	33
3.7	<i>Mahalanobis Distance</i>	35
3.7.1	Hitung Kovarian Matrix	36
3.7.2	Hitung <i>Mahalanobis Distance</i>	38
3.7.3	Hasil <i>Mahalanobis Distance</i>	39
3.8	Pengujian	39
3.9	Kesimpulan.....	40
BAB IV HASIL DAN PEMBAHASAN		41
4.1	Data Penelitian	41
4.2	Pembahasan	41
4.2.1	Tahap Text Preprocessing	41
4.2.2	Tahap Pembobotan.....	44
4.2.3	Tahap Mahalanobis Distance	48
4.3	Pengujian	52

4.3.1	Pengujian <i>Mahalanobis</i>	52
4.3.2	Pengujian <i>Threshold</i>	59
4.4	Kesimpulan Pengujian.....	67
BAB V KESIMPULAN DAN SARAN.....		69
5.1	Kesimpulan.....	69
5.2	Saran	69
DAFTAR PUSTAKA		71
DATA PENELITIAN		75
BIODATA PENULIS		86
KARTU BIMBINGAN.....		87
SURAT PERNYATAAN TIDAK PLAGIAT		88