

# Menentukan Tingkat Kemiripan Judul Mahasiswa Fakultas Keguruan dan Ilmu Pendidikan Unismuh Makassar Menggunakan Metode Cosine Similarity

Haedir<sup>\*1</sup>, Lukman<sup>1</sup>, Titin Wahyuni<sup>1</sup>

<sup>1</sup>Program Studi Informatika, Fakultas Teknik, Universitas Muhammadiyah Makassar  
e-mail koresponden: [haedirbaba1705@gmail.com](mailto:haedirbaba1705@gmail.com)

Received: Agustus 01,2025; Accepted: Agustus 01, 2025; Published: 30 September, 2025

## Abstrak

*Plagiarisme dan duplikasi judul skripsi menjadi tantangan serius dalam menjaga orisinalitas penelitian mahasiswa di Fakultas Keguruan dan Ilmu Pendidikan (FKIP) Universitas Muhammadiyah Makassar. Penelitian ini bertujuan mengimplementasikan metode cosine similarity untuk mendeteksi kemiripan judul skripsi dan mengevaluasi kinerjanya menggunakan metrik standar. Data penelitian meliputi 1.000 judul skripsi yang diproses melalui tahapan preprocessing, ekstraksi fitur TF-IDF, perhitungan cosine similarity, hingga evaluasi model. Hasil penelitian menunjukkan sistem mampu mendeteksi kemiripan dengan akurasi 87,33%, presisi 100%, recall 58,70%, dan F1-score 73,97%. Presisi sempurna mengindikasikan sistem sangat andal dalam mengidentifikasi judul mirip tanpa kesalahan positif palsu. Namun, recall yang masih rendah menunjukkan beberapa judul mirip belum terdeteksi optimal. Penelitian ini memberikan kontribusi praktis sebagai alat bantu verifikasi keaslian judul skripsi dan mendorong pengembangan sistem deteksi kemiripan yang lebih sensitif di masa mendatang.*

**Kata kunci:** Cosine similarity, TF-IDF, Kemiripan Judul, Deteksi Plagiarisme, Natural Language Processing.

## Abstract

*Plagiarism and duplicate thesis titles pose serious challenges to maintaining research originality among students at the Faculty of Teacher Training and Education (FKIP), Universitas Muhammadiyah Makassar. This study aims to implement the cosine similarity method to detect thesis title similarity and evaluate its performance using standard metrics. The research data comprised 1,000 thesis titles processed through preprocessing stages, TF-IDF feature extraction, cosine similarity calculation, and model evaluation. Results show the system can detect similarity with 87.33% accuracy, 100% precision, 58.70% recall, and 73.97% F1-score. Perfect precision indicates the system is highly reliable in identifying similar titles without false positives. However, the relatively low recall indicates that some similar titles remain undetected. This research provides practical contributions as a tool for verifying the authenticity of thesis titles and encourages the development of more sensitive similarity-detection systems in the future.*

**Keyword:** Cosine similarity, TF-IDF, Title Similarity, Plagiarism Detection, Natural Language Processing.

## 1. Pendahuluan

Dalam era digital yang serba cepat seperti sekarang, teknologi informasi telah menjadi bagian penting dalam dunia akademik. Fakultas Keguruan dan Ilmu Pendidikan (FKIP) di Universitas Muhammadiyah (Unismuh) Makassar menghadapi tantangan dalam mengelola dan menilai kesamaan judul skripsi mahasiswa. Keunikan dan orisinalitas judul skripsi sangatlah penting untuk mencegah plagiarisme dan memastikan keragaman topik penelitian [1], [2].

Setiap tahunnya, FKIP Unismuh Makassar menerima ratusan proposal skripsi dari mahasiswa yang mencakup berbagai program studi. Proses verifikasi manual terhadap keunikan judul-judul tersebut memerlukan waktu dan tenaga yang tidak sedikit, terutama ketika harus membandingkan dengan database judul-judul terdahulu yang terus bertambah. Tanpa sistem yang terstruktur, risiko terjadinya duplikasi atau kemiripan topik penelitian yang berlebihan menjadi sulit

dikendalikan, yang pada akhirnya dapat berdampak pada kualitas dan kontribusi keilmuan yang dihasilkan.

Seiring berkembangnya teknologi kecerdasan buatan, terutama di bidang pengolahan bahasa alami atau *Natural Language Processing* (NLP), berbagai metode analisis teks mulai diterapkan. NLP memungkinkan komputer memahami dan mengolah bahasa manusia, yang menjadi dasar untuk analisis kemiripan teks [3].

Dalam konteks bahasa Indonesia, penerapan metode NLP untuk deteksi kemiripan teks menghadapi tantangan tersendiri dibandingkan dengan bahasa Inggris. Karakteristik morfologi bahasa Indonesia yang kompleks, termasuk penggunaan afiks (awalan, sisipan, akhiran) dan reduplikasi, memerlukan pendekatan *preprocessing* yang lebih hati-hati. Selain itu, variasi penulisan istilah akademik dan terminologi pendidikan dalam judul-judul skripsi menambah kompleksitas dalam proses analisis kemiripan [4].

Plagiarisme merupakan salah satu masalah serius dalam dunia akademik [5]. Untuk mengatasi masalah ini, diperlukan penerapan dan pemahaman mendalam terhadap metode yang dapat menganalisis kemiripan teks, guna mengarahkan pada penggunaan yang lebih baik. Metode *cosine similarity* adalah salah satu pendekatan yang umum digunakan untuk menghitung tingkat kesamaan antara dua item teks, seperti judul skripsi [6], [7]. Pendekatan ini bekerja dengan cara membandingkan representasi numerik dari item-item tersebut. Dengan adanya pembobotan TF-IDF (*Term Frequency - Inverse Document Frequency*), metode *cosine similarity* dapat memproses kata secara lebih optimal [8], [9].

Berbagai penelitian terdahulu telah menunjukkan efektivitas *cosine similarity* dalam berbagai aplikasi deteksi kemiripan teks, mulai dari sistem rekomendasi, information retrieval, hingga deteksi plagiarisme dokumen [10], [11], [12], [13]. Namun, penerapannya secara spesifik pada domain judul skripsi pendidikan di Indonesia masih relatif terbatas. Judul skripsi memiliki karakteristik unik berupa teks yang relatif pendek namun padat informasi, yang berbeda dengan dokumen panjang seperti artikel atau konten web yang menjadi fokus sebagian besar penelitian sebelumnya [6], [14], [15]. Oleh karena itu, perlu dilakukan kajian mendalam mengenai bagaimana metode ini bekerja dalam konteks data yang spesifik ini.

Penelitian ini bertujuan untuk mengimplementasikan metode *cosine similarity* dan mengetahui kinerjanya dalam mendeteksi kemiripan judul skripsi mahasiswa di FKIP Unismuh Makassar. Fokus utama adalah pada pemahaman sejauh mana metode ini mampu memberikan hasil yang akurat dan relevan dalam konteks dataset yang ada, sehingga dapat menjadi panduan untuk penerapannya yang lebih baik. Penelitian ini akan melibatkan tahapan implementasi metode, pengujian dengan data riil, dan evaluasi hasil menggunakan metrik standar seperti akurasi, presisi, *recall*, dan *F1-score*.

Dataset yang digunakan dalam penelitian ini merupakan kumpulan judul skripsi mahasiswa FKIP Unismuh Makassar dari berbagai program studi dalam rentang waktu tertentu. Keberagaman topik penelitian yang mencakup berbagai bidang Pendidikan, mulai dari pendidikan matematika, bahasa, IPA, hingga pendidikan guru sekolah dasar, memberikan variasi data yang cukup representatif untuk menguji *robustness* metode *cosine similarity*. Melalui analisis mendalam terhadap hasil pengukuran kemiripan, penelitian ini akan mengidentifikasi pola-pola kemiripan yang sering muncul serta *threshold* nilai *similarity* yang paling optimal untuk menandai judul-judul yang perlu ditinjau lebih lanjut.

Penggunaan metode *cosine similarity* untuk penelitian ini didasarkan pada relevansinya dalam studi kemiripan teks. Penelitian ini juga akan mempertimbangkan keterbatasan *cosine similarity* dalam mengenali sinonim sebagai bagian dari diskusi hasil evaluasi [14].

Dengan mengimplementasikan dan mengetahui kinerja metode *cosine similarity*, diharapkan penelitian ini dapat memberikan pemahaman konkret mengenai kemampuannya dalam membantu proses verifikasi keunikan judul skripsi di Universitas Muhammadiyah (Unismuh) Makassar. Penelitian ini bukan merupakan pengembangan sistem machine learning yang melibatkan pelatihan model prediktif kompleks, melainkan sebuah studi implementatif dan evaluatif terhadap sebuah metode pengukuran kemiripan teks.

## 2. Metode Penelitian

### 2.1. Tempat dan Waktu Penelitian

Tempat Penelitian adalah suatu tempat atau objek yang akan dilakukan suatu penelitian. Penentuan lokasi penelitian merupakan langkah penting dalam proses penelitian karena

memudahkan peneliti untuk melakukan penelitian. Lokasi penelitian yang dipilih oleh penulis adalah di Universitas Muhammadiyah Makassar. Tempat atau wilayah tersebut dipilih oleh penulis dengan alasan karena tempat penelitian atau lebih tepatnya Fakultas Keguruan dan Ilmu Pendidikan Universitas Muhammadiyah Makassar pada proses pembuatan jurnal ilmiah masih dilakukan dengan cara manual sehingga seringkali menguras banyak waktu dan tenaga mahasiswa, oleh karena itu peneliti memilih tempat penelitian tersebut dikarenakan masalah yang diangkat oleh peneliti sehaluan dengan tempat tersebut. Waktu penelitian ini akan dilakukan dalam jangka waktu kurang lebih 2 bulan, yaitu dimulai pada bulan Januari 2025 sampai dengan Maret 2025.

## 2.2. Metode Perancangan Sistem

Penelitian ini dirancang mengikuti tahapan implementasi dan evaluasi kemiripan judul skripsi menggunakan pendekatan *cosine similarity* terhadap representasi vektor TF-IDF, dengan langkah-langkah sebagai berikut:

### 1. Pengumpulan

Langkah pertama adalah mengumpulkan data berupa judul-judul skripsi dari mahasiswa FKIP Universitas Muhammadiyah Makassar. Sebagian data digunakan sebagai data latih, dan sebagian lainnya disiapkan sebagai data uji (*test set*) yang akan divalidasi menggunakan label manual atau anotasi dari dosen sebagai ground truth.

### 2. Preprocessing Data

Data judul yang telah dikumpulkan kemudian diproses secara teks dengan beberapa tahap:

- Tokenisasi, yaitu memecah teks judul menjadi unit kata.
- Stop Words Removal*, untuk menghapus kata-kata umum yang tidak relevan (seperti "yang", "dalam", "dari").
- Stemming*, yaitu mengembalikan kata ke bentuk dasar agar kata-kata turunan dihitung sebagai satu entitas yang sama.

### 3. Ekstraksi Fitur

Setelah *preprocessing*, setiap judul dikonversi menjadi representasi numerik menggunakan teknik TF-IDF (*Term Frequency-Inverse Document Frequency*). Bobot TF-IDF menunjukkan tingkat kepentingan suatu kata dalam sebuah judul relatif terhadap keseluruhan koleksi dokumen.

### 4. Cosine similarity

Langkah ini menghitung skor *cosine similarity* antara dua judul skripsi berdasarkan vektor TF-IDF mereka. *Cosine similarity* mengukur derajat kesamaan berdasarkan sudut antar vektor, dengan nilai 1 menunjukkan kemiripan sempurna dan 0 menunjukkan tidak ada kemiripan.

### 5. Evaluasi Model

Model dievaluasi menggunakan data uji. Hasil *cosine similarity* dibandingkan dengan label kemiripan manual untuk mengukur akurasi sistem. Validasi ini penting untuk mengetahui apakah sistem mendeteksi kemiripan sebagaimana yang dinilai oleh pakar.

### 6. Deteksi

Berdasarkan nilai *cosine similarity* dan ambang batas (*threshold*) yang telah ditentukan, sistem akan mengklasifikasikan setiap pasangan judul sebagai "mirip" atau "tidak mirip".

### 7. Interpretasi Hasil

Tahap akhir adalah menyajikan hasil deteksi dalam bentuk *output* yang mudah dipahami. Sistem akan menampilkan skor kemiripan dan status deteksi (mirip/tidak mirip) sebagai dasar pengambilan keputusan lebih lanjut oleh dosen pembimbing atau tim akademik.

## 2.3. Metode Pengujian Sistem

Pengujian sistem dilakukan untuk memastikan bahwa sistem yang dibangun sudah sesuai dengan analisis dan tujuan penelitian. Berikut adalah langkah-langkah pengujian sistem:

### 1. Persiapan Data Uji (*Ground Truth*):

- Sebuah subset dari dataset judul akan dipilih sebagai data uji.
- Pasangan-pasangan judul dalam data uji ini akan dinilai secara manual oleh ahli (misalnya dosen pembimbing) untuk menentukan apakah mereka benar-benar mirip atau tidak. Hasil penilaian manual ini akan menjadi ground truth.

### 2. Skenario Pengujian:

- Setiap judul dalam data uji akan dibandingkan dengan judul-judul lain (atau judul referensi) menggunakan metode *cosine similarity* yang telah diimplementasikan.
- Berdasarkan skor *cosine similarity* dan *threshold* yang telah ditentukan, sistem akan memprediksi apakah pasangan judul tersebut mirip atau tidak.

3. Pengumpulan Hasil Prediksi: Hasil prediksi dari metode (mirip/tidak mirip) akan dicatat untuk setiap pasangan dalam data uji.

## 2.4. Metode Analisis

1. Perhitungan *Similarity*

Langkah pertama dalam proses evaluasi adalah menghitung kemiripan (*similarity*) antar judul skripsi atau teks menggunakan metode *cosine similarity*. *Cosine similarity* mengukur seberapa mirip dua vektor dalam ruang multidimensi, dalam hal ini representasi teks atau fitur. Nilai *similarity* berkisar antara 0 hingga 1, di mana nilai mendekati 1 menunjukkan kemiripan yang sangat tinggi. Dalam konteks ini, setiap pasangan judul skripsi dibandingkan untuk melihat kemiripan judul skripsi berdasarkan nilai *cosine similarity*.

2. Analisis Hasil Metrik

Setelah nilai *similarity* dihitung, langkah berikutnya adalah membuat label ground truth. Ground truth adalah data pembanding (label sebenarnya) yang digunakan untuk mengevaluasi akurasi hasil prediksi. Dalam proses ini, digunakan ambang batas (*threshold*) tertentu, misalnya 0.5. Jika nilai *similarity* antara judul skripsi sama dengan atau lebih besar dari *threshold* ini, maka pasangan tersebut diberi label "similar". Sebaliknya, jika nilainya di bawah *threshold*, pasangan tersebut dianggap "tidak similar". *Ground truth* ini menjadi referensi untuk mengukur kinerja sistem prediksi.

3. Diskusi dan Interpretasi

Untuk mengetahui sensitivitas sistem terhadap tingkat kemiripan yang berbeda, dilakukan pengujian dengan menggunakan beberapa nilai *threshold*, misalnya 0.6, 0.7, dan 0.8. Setiap nilai *threshold* ini menghasilkan prediksi yang berbeda. Semakin tinggi *threshold*, semakin ketat sistem dalam menganggap dua judul sebagai "mirip". Dengan mencoba berbagai ambang batas, peneliti dapat mengevaluasi performa sistem pada skenario yang berbeda dan memilih *threshold* yang paling sesuai dengan kebutuhan aplikasi.

4. Perhitungan *Metrics*

Tahap terakhir adalah mengevaluasi performa sistem prediksi dengan membandingkan hasil prediksi terhadap ground truth. Pengukuran dilakukan dengan menghitung metrik evaluasi umum seperti *accuracy* (tingkat ketepatan klasifikasi), *precision* (ketepatan prediksi positif), *recall* (kemampuan mendeteksi seluruh data positif), dan *f1-score* (rata-rata harmonis antara *precision* dan *recall*). Dengan metrik-metrik ini, performa sistem dapat dianalisis secara komprehensif untuk menentukan seberapa baik sistem membedakan judul yang mirip dan tidak mirip.

## 3. Hasil dan Pembahasan

### 3.1. Hasil Implementasi

1. Pengumpulan Data

Langkah pertama dalam penelitian ini adalah memperoleh data judul skripsi dari tabel *judul\_skripsi* pada database MySQL. Gambar 1 menyajikan data yang berhasil dikumpulkan terdiri atas dua atribut utama, yakni nomor identitas dan judul skripsi, dengan total sebanyak 1000 judul. Selama proses ekstraksi, dilakukan pengecekan agar setiap baris data benar-benar sesuai dan dapat digunakan untuk proses analisis selanjutnya. Data ini nantinya akan menjadi sumber utama yang diolah dalam tahap pra-pemrosesan serta pelatihan model untuk mengukur tingkat kemiripan antar judul skripsi.

1	penulis	judul_skripsi
2	AHMAD ZULFIKAR	Pengaruh Tingkat Pendidikan Orang Tua Terhadap Pembentukan Kepribadian Anak (Studi Komunikasi Dalam Keluarga Pada Lingkungan Calie Kabupaten Sinjai)
3	A FAHUL HAQ	BUDAYA KONSUMTIF MASYARAKAT SINJAI (STUDI KASUS PADA LAPANGAN SINJAI BERSATU)
4	A FITRIANI	Pengaruh Media Gambar Terhadap Minat Baca Anak Usia Dini di TK Insan Cemerlang Manuruk Makassar
5	A INDASHARY	TINDAK TUTUR PAPASENG MASYARAKAT KAJURA KABUPATEN BONE
6	A INDASHARY	TINDAK TUTUR PAPASENG MASYARAKAT KAJURA KABUPATEN BONE
7	A ISMAYANTI	PERANAN PENDIDIKAN PKN DALAM MENUMBUHKAN SKAP NASIONALISME SISWA SMA NASIONAL MAKASSAR
8	A MULI HADIRINTO	EFEKTIVITAS PENYULUHAN PERTANIAN TERHADAP PERUBAHAN SOSIAL EKONOMI MASYARAKAT KAJURA KABUPATEN BONE
9	A NUR INAYAH RAHMAD	PENGARUH PENGGUNAAN METODE DISCOVERY BERBASIS MEDIA REALITA TERHADAP HASIL BELAJAR IPA KONSEP SIFAT DAN PERUBAHAN WILUD BENDA SISWA KELAS IV SD INPRES 8 KALIMBANG KECAMATAN BANTAENG KABUPATEN BANTAENG
10	A NURDAWANI	Pengaruh Penggunaan Bahasa Indonesia Terhadap Pembentukan Pola Pikir Masyarakat Siswa Kelas V SD Inpres Bontomanta Kota Makassar
11	A RESKI AMALIA USMAN	PENERAPAN METODE SAINTIFIK BERBASIS LITERASI SAINT TERHADAP KETERAMPILAN PROSES SAINS PESERTA DIDIK MA MUALLIMIN MUHAMMADIYAH MAKASSAR
12	A RESKANTI ARDI	Efektivitas Pembelajaran Matematika Melalui Penerapan Pendekatan Realistic Mathematics Education (RME) pada Siswa Kelas VIII SMP Unisuruh Makassar
13	A RUSGANDI SURYA NINGRAT	ENGKEMBANGAN USAHA DAN PROBLEMA MASYARAKAT DALAM KAWASAN HUTAN KECAMATAN SIMBORO KABUPATEN MAMUUJU SULAWESI BARAT
14	A SERLY ANGGRANI	PENGARUH METODE KOOPERATIF Tipe THINK PAIR SHARE (TPS) TERHADAP HASIL BELAJAR FISIKA PADA PESERTA DIDIK MAN BANTAENG
15	A SYA AFIRA	THE USE OF TALKING CHIPS METHOD IN IMPROVING STUDENTS SPEAKING SKILL OF SMP 39 BULUKUMBA
16	A SAIFUL SAHRIN	Perilaku Sosial Siswa dalam Menggunakan Media Sosial Facebook
17	A UMUL HARA	PENGARUH PENERAPAN METODE STORYTELLING TERHADAP HASIL BELAJAR KETERAMPILAN BERBICARA PADA MATA PELAJARAN BAHASA INDONESIA SISWA KELAS V SD INPRES PULAEWNG KECAMATAN EREMERA KABUPATEN BANTAENG
18	A WAHYUDI	KULTUR PESANTREN DALAM PEMBINAAN KARAKTER SANTRI DI PONDOK PESANTREN DARUL ARQAM MUHAMMADIYAH GOMBONG MAKASSAR
19	A YUNIARTI DIAN PRATIWI	THE EFFECT OF MNEMONIC TECHNIQUE WITH RESPECT TO LEARNERS' VOCABULARY IN TEACHING VOCABULARY TO LEARNERS AT SMA 11 WATAMPONE (A PRE-EXPERIMENTAL RESEARCH)
20	AAKIRA WIBOWO	EKSISTENSI KOMUNITAS LITERASI UNIVERSITAS MUHAMMADIYAH MAKASSAR (SUATU KAJIAN RELASI KUASA MICHEL FOUCAULT)
21	A ASRIANI	Efektivitas Pembelajaran Matematika Melalui Pendekatan Kontekstual pada Siswa Kelas VII-2 SMP Negeri 4 Bajeng Kabupaten Gowa
22	A HIKMA WARDANI	Penerapan Metode Empiris pada Pembelajaran Fisika Terhadap Pengetahuan Hasil Belajar Peserta Didik Kelas XI SMA Negeri 13 Luwu
23	A HIKMA WARDANI	Penerapan Metode Empiris pada Pembelajaran Fisika Terhadap Pengetahuan Hasil Belajar Peserta Didik Kelas XI SMA Negeri 13 Luwu
24	A HILATUL MUSLIMAH	EFEKTIVITAS PEMBELAJARAN MATEMATIKA DENGAN PENERAPAN METODE KOMIK PADA SISWA KELAS X IKU SMA MUHAMMADIYAH GALESONG
25	A INDRIYANTI	EFEKTIVITAS PEMBELAJARAN MATEMATIKA DENGAN MENGGUNAKAN MULAIKU (MEDIA INTERAKTIF DAN INOVATIF) PADA SISWA KELAS VII SMP NEGERI 2 PANGAJENE
26	A MUHAMMAD	EFEKTIVITAS PEMBELAJARAN MATEMATIKA MELALUI PENGGUNAAN METODE KOOPERATIF Tipe NUMBERED HEADS TOGETHER (NHT) PADA SISWA KELAS XI IPA SMA MUHAMMADIYAH WILAYAH MAKASSAR
27	A MUSDAUFAH	Desripsi Pembelajaran Matematika melalui Model Quantum Teaching pada Siswa Kelas IV VII SMP Negeri 14 Pallanga Kabupaten Gowa
28	A NURASIHAFI HASYIM	PENGARUH PENERAPAN METODE STRUKTURAL ANALITIK SINTETIK (SAS) TERHADAP KEMAMPUAN MEMBACA PEMERILAIAN SISWA KELAS I SD NEGERI 101 KASUZO KECAMATAN BONTOBARI KABUPATEN BULUKUMBA
29	A NURQISHAF HASYIM	PENGARUH PENERAPAN METODE STRUKTURAL ANALITIK SINTETIK (SAS) TERHADAP KEMAMPUAN MEMBACA PEMERILAIAN SISWA KELAS I SD NEGERI 101 KASUZO KECAMATAN BONTOBARI KABUPATEN BULUKUMBA
30	A NURFAIRIANA	EFEKTIVITAS PEMBELAJARAN MATEMATIKA MELALUI PENERAPAN MODEL REPLICATIVE TEACHING SETTING KOOPERATIF PADA SISWA KELAS VIII SMP NEGERI 15 BULUKUMBA
31	A NURINORA V	ANALISIS PROGRAM PENDIDIKAN KARAKTER PADA MATA KELAJUAN PENDIDIKAN AGAMA ISLAM (PAI) DI SD NEGERI UNGGULAN BONTONTOANE NGEKALI KABUPATEN GOWA
32	A ROSIDINA AMIRI	MAXIMA SOSIAL SIMBOLISME TALARAN TERHADAP IMPLEMENTASI MOTIVASI INDUDE SINGLE PARENT (STUDI KASUS DESA LABAJA KECAMATAN CENRANA KABUPATEN BONE)
33	A SUARYANINGSIH ARIYA	Perbedaan Hasil Belajar Matematika antara Model Kooperatif Tipe TGT dan NHT pada Kelas VIII SMP Somba Opu
34	A RATNA PRATINI PUTRI	HUBUNGAN POLA ASIH ORANG TUA DENGAN KEMANDIRIAN ANAK USAIA 5-6 TAHUN DI KELAS XA KABUPATEN KEPULAUAN SELAYAR
35	A WAHYUNI RAMZANA	ANALISIS KESALAHAN MENYELESAIKAN SOAL: CERTA MATERI GEOMETRI BERDASARAN TEORI NEWMAN PADA KELAS VIII SMP NEGERI 1 SUNGGUMINASA
36	AAN INDRAYADI	STUDENTS' ABILITY IN IDENTIFYING INDEPENDENT CLAUSE AND DEPENDENT CLAUSE IN COMPLEX SENTENCE
37	ABD QADIR JAELANI	Efektivitas Pembelajaran Matematika Melalui Model Kooperatif Tipe Talking Stick pada Siswa Kelas VII SMP Negeri 1 Salome Kabupaten Bone
38	ABD RAHMAN	Hubungan Pengelolaan Kela Terhadap Motivasi Belajar Murid SDN Genteng Bajeng Barat Kabupaten Gowa
39	ABD RAHMAN	PENINGKATAN KEMAMPUAN MENULIS KARANGAN NARASI MENGUNAKAN STRATEGI AKTIVITAS TERBIMBING MURID KELAS V SD INPRES 10 KAPASA KABUPATEN JENEPOPO
40	ABD RAHMAN	PENINGKATAN KEMAMPUAN MENULIS KARANGAN NARASI MENGUNAKAN STRATEGI AKTIVITAS TERBIMBING MURID KELAS V SD INPRES 10 KAPASA KABUPATEN JENEPOPO

Gambar 1. Judul skripsi hasil ekstraksi dari database

## 2. Preprocessing Data

Pra-pemrosesan data merupakan tahapan awal untuk memastikan data yang digunakan sudah bersih dan seragam. Salah satu proses utama dalam tahap ini adalah mengkonversi seluruh teks judul skripsi menjadi huruf kecil seperti yang ditunjukkan pada Gambar 2 dan 3. Cara ini diterapkan untuk menghindari masalah duplikasi akibat perbedaan penulisan huruf kapital, sehingga proses analisis dapat berjalan lebih konsisten dan hasil yang diperoleh menjadi lebih valid. Tahapan ini termasuk dalam proses *Natural Language Processing* (NLP) yang bertujuan meningkatkan kualitas data sebelum dianalisis lebih lanjut. Melalui tahapan ini, seluruh judul skripsi yang digunakan dalam penelitian telah memiliki format yang seragam dan siap untuk dianalisis pada tahap selanjutnya, yaitu penghitungan tingkat kemiripan.

20	PENERAPAN MODEL INKLUZY TERBIMBING MENINGKATKAN HASIL BELAJAR IPA KELAS V SD LEMBAYA KECAMATAN TOMPOLU KABUPATEN GOWA
21	PENINGKATAN HASIL BELAJAR MATEMATIKA PENDEKATAN REALISTIK MATEMATIKA (RME) MURID KELAS IV SON NO. CENTRE PATTALLASSANG KECAMATAN PATTALLASSANG KABUPATEN TAKALAR
22	PENINGKATAN HASIL BELAJAR IPA DENGAN MODEL PEMBELAJARAN OUTDOOR LEARNING MURID KELAS V SD NEGERI NO. MAERO KECAMATAN BONTORAMBA KABUPATEN JENEPOPO
23	PENINGKATAN KEMAMPUAN MENULIS KARANGAN DESKRIPSI DENGAN TEKNIK MIND MAP MURID KELAS II SON INPRES BONTOPAJUA KECAMATAN GALESONG UTARA KABUPATEN TAKALAR
24	KESANTUNAN BERBAHASA INDONESIA SISWA DAN GURU DALAM PROSES KEGIATAN PEMBELAJARAN DARING PADA MASA PANDEMI SISWA KELAS VII SMPN SATAP PUNUGA KABUPATEN TAKALAR
25	IMPLEMENTASI KEBIJAKAN SEKOLAH RUIJUKAN SMP NEGERI SUNGGUMINASA KABUPATEN GOWA
26	EFEKTIVITAS PENERAPAN MODEL PEMBELAJARAN PROBLEM BASED LEARNING TERHADAP KEMAMPUAN BERPIKIR KRITIS SISWA MATA BELAJAR PPKN KELAS V SD NEGERI DENA
27	PENGEMBANGAN MEDIA PEMBELAJARAN FISIKA BERBASIS ADOBE AIR FOR ANDROID TERHADAP DUA DUKUNG PEMBELAJARAN PESERTA DIDIK
28	PENERAPAN MODEL PEMBELAJARAN THINK PAIR SHARE (TPS) UNTUK MENINGKATKAN HASIL BELAJAR IPA SISWA KELAS V SD NEGERI PANCIRI KECAMATAN BAENG KABUPATEN GOWA
29	PENGARUH PENGGUNAAN MODEL PROBLEM CENTERED LEARNING (PCL) TERHADAP KEMAMPUAN MENULIS NARASI SISWA KELAS V SD SALOBUNDANG KECAMATAN BONTOTIRO KABUPATEN BULUKUMBA
30	PENGARUH KEPERIMIMPINAN GURU TERHADAP MOTIVASI BELAJAR PESERTA DIDIK SD NEGERI BANUA KABUPATEN MAJENE
31	ANALISIS TINGKAT KESULITAN SISWA MENULIS CERTA PADA MATA PELAJARAN BAHASA INDONESIA KELAS IV SD MANNURUKI
32	ANALISIS KEMAMPUAN PEMAHAMAN KONSEP MATEMATIKA MATERI GARIS DAN SUDUT DITINJAU DARI GAYA KOGNITIF SISWA KELAS VII SMPN PALLANGGA
33	WHATSAPP SEBAGAI MEDIA PEMBELAJARAN PADA MASA PANDEMI COVID-19 DI SDN TIMBULA KECAMATAN BONTOTIRO KABUPATEN BULUKUMBA
34	ANALISIS KETERAMPILAN BERPIKIR KRITIS SISWA DALAM PEMBELAJARAN FISIKA DARING DI SMA MUHAMMADIYAH MAKASSAR
35	PENGARUH PENGGUNAAN MEDIA PEMBELAJARAN BERBASIS POWERPOINT SPARKOL VIDEOSCRIBE TERHADAP MINAT BELAJAR SISWA MATA PELAJARAN BIOLOGI KELAS VII SMP PGRI BONTORAMBA KECAMATAN BONTORAMBA
36	TIPOLOGI SOLIDARITAS SOSIAL PETAINI PENGGARAP DI KABUPATEN MARO DITINJAU DARI TEORI PERTUKARAN SOSIAL
37	ANALISIS PENGGUNAAN MEDIA PEMBELAJARAN BERBASIS POWERPOINT SPARKOL VIDEOSCRIBE TERHADAP MINAT BELAJAR SISWA MATA PELAJARAN BIOLOGI KELAS VII SMP PGRI BONTORAMBA
38	PROSES BERPIKIR SISWA DALAM PEMECAHAN MASALAH MATEMATIKA BERDASARKAN GAYA KOGNITIF SISWA KELAS VIII SMP NEGERI BAENG
39	PENGARUH PENGGUNAAN MEDIA BIG BOOK TERHADAP HASIL BELAJAR IPA SISWA KELAS IVB SD INPRES PAKU KECAMATAN PALLANGGA KABUPATEN GOWA
40	PENERAPAN SIKLUS BELAJAR KARPLUS TERHADAP KETERAMPILAN PROSES SAINS PESERTA DIDIK MADRASAH ALIYAH BONTOMARANNU
41	DESKRIPSI KEMAMPUAN SISWA MEMECAHKAN MASALAH GEOMETRI BERDASARKAN TEORI VAN HIELE KELAS VIII SMP MAKASSAR
42	PENGEMBANGAN INSTRUMEN BERBASIS KETERAMPILAN BERPIKIR KREATIF PESERTA DIDIK PADA MATERI FLUIDA STATIS SMA NEGERI GOWA
43	ANALISIS KEMAMPUAN REPRESENTASI MATEMATIS MATERI HIMPUNAN DITINJAU DARI GAYA KOGNITIF SISWA KELAS VII SMP NEGERI MAKASSAR
44	DESKRIPSI KEMAMPUAN BERPIKIR KRITIS MATEMATIS SISWA DALAM MENYELESAIKAN SOAL MATERI RELASI DAN FUNGSI DITINJAU DARI PERBEDAAN GENDER SISWA KELAS VIII SMP MAKASSAR
45	EFEKTIVITAS PEMBELAJARAN MATEMATIKA DENGAN PENDEKATAN ELPSS PADA PESERTA DIDIK KELAS VIII A SMP NEGERI GALESONG UTRA

Gambar 2. Data sebelum preprocessing

20	penerapan model inkui terbimbing meningkatkan hasil belajar ipa kelas v sdn lembaya kecamatan tombolebu kabupaten gowa
21	peningkatan hasil belajar matematika pendekatan realistik matematika (rme) murid kelas iv sdn no. centre pattallassang kecamatan pattallassang kabupaten takalar
22	peningkatan hasil belajar ipa dengan model pembelajaran outdoor learning murid kelas v sd negeri no. maero kecamatan bontorumba kabupaten jeneponto
23	peningkatan kemampuan menulis karangan deskripsi dengan teknik mind map murid kelas ii sdn inpres bontopaja kecamatan galesong utara kabupaten takalar
24	kesantran berbahasa indonesia siswa dan guru dalam proses kegiatan pembelajaran daring pada masa pandemi siswa kelas vii smpn satap punaga kabupaten takalar
25	implementasi kebijakan sekolah rujukan smp negeri sungsungan kabupaten gowa
26	efektivitas penerapan model pembelajaran problem based learning terhadap kemampuan berpikir kritis siswa mata pelajaran ppkn kelas v sd negeri dena
27	pengembangan media pembelajaran fisika berbasis adobe air for android terhadap daya dukung pembelajaran peserta didik
28	penerapan model pembelajaran think pair share (tps) untuk meningkatkan hasil belajar ipa siswa kelas v sd negeri panciro kecamatan bajeng kabupaten gowa
29	pengaruh penggunaan model pembelajaran learning (pcl) terhadap kemampuan menulis narasi siswa kelas v sdn salobundang kecamatan bontotiro kabupaten bulukumba
30	pengaruh kelemparan guru terhadap motivasi belajar peserta didik sd negeri banu kabupaten majene
31	analisis tingkat kesulitan siswa menulis cerita pada mata pelajaran bahasa indonesia kelas iv sdn manuruk
32	analisis kemampuan pemahaman konsep matematika materi garis dan sudut ditinjau dari gaya kognitif siswa kelas vii smpn palangga
33	whatsapp sebagai media pembelajaran pada masa pandemi covid-19 di sdn timbulu kecamatan bontotiro kabupaten bulukumba
34	analisis keterampilan berpikir kritis siswa dalam pembelajaran fisika daring di sma muhammadiyah makassar
35	pengaruh penggunaan media kotak huruf terhadap kemampuan membaca permulaan siswa kelas i sdn karanglo kecamatan tombolebu kabupaten gowa
36	tipologi solidaritas sosial petani pengaruh di kabupaten maro ditinjau dari teori pertukaran sosial
37	analisis penggunaan media pembelajaran berbasis powerpoint sparkol videoscribe terhadap minat belajar siswa mata pelajaran biologi kelas vii smp pgri bontorumba kecamatan bontorumba
38	proses berpikir siswa dalam pemecahan masalah matematika berdasarkan gaya kognitif siswa kelas viii smp negeri bajeng
39	pengaruh penggunaan media big book terhadap hasil belajar ipa siswa kelas ivb sd inpres paku kecamatan pallangga kabupaten gowa
40	peneparan siklus belajar karplus terhadap keterampilan proses sains peserta didik madrasah aliyah bontomarannu
41	deskripsi kemampuan siswa memecahkan masalah geometri berdasarkan teori von hiele kelas viii a smpn makassar
42	pengembangan instrumen berbasis keterampilan berpikir kreatif peserta didik pada materi fluida statis sma negeri gowa
43	analisis kemampuan representasi matematis materi himpunan ditinjau dari gaya kognitif siswa kelas viii smp negeri makassar
44	deskripsi kemampuan berpikir kritis matematis siswa dalam menyelesaikan soal materi relasi dan fungsi ditinjau dari perbedaan gender siswa kelas viii smpn makassar
45	efektivitas pembelajaran matematika dengan pendekatan elpsa pada peserta didik kelas viii a smp negeri galesong utara

Gambar 3. Data sesudah preprocessing

### 3. Ekstraksi Fitur (Representasi TF-IDF).

Setelah data judul skripsi selesai dibersihkan melalui proses *preprocessing*, tahap selanjutnya adalah melakukan ekstraksi fitur menggunakan metode TF-IDF (*Term Frequency - Inverse Document Frequency*) dengan skrip yang ditampilkan pada Gambar 4. Tujuan dari tahap ini adalah mengubah kumpulan kata-kata pada judul skripsi menjadi bentuk angka atau nilai bobot, agar bisa dihitung oleh komputer. Secara sederhana, TF-IDF akan melihat seberapa sering sebuah kata muncul dalam satu judul (*Term Frequency*), namun juga mempertimbangkan apakah kata tersebut sering muncul di banyak judul lain (*inverse document frequency*). Dengan begitu, kata-kata yang sering muncul di satu judul tetapi jarang muncul di judul lain akan memiliki bobot yang lebih tinggi, karena dianggap lebih penting. Sebaliknya, kata-kata yang muncul hampir di semua judul (misalnya "analisis", "pengaruh") akan memiliki bobot lebih rendah. Hasil dari proses ini adalah sebuah vektor angka yang mewakili isi dari setiap judul skripsi. Vektor inilah yang nantinya digunakan untuk mengukur kemiripan antar judul pada tahap selanjutnya.

```
def extract_features(self):
    if not hasattr(self, 'processed_titles'):
        self.log("No processed data available")
        return

    self.tfidf_matrix = self.tfidf_vectorizer.fit_transform(self.
    processed_titles)
    self.log(f"Feature extraction completed. TF-IDF matrix shape: {self.
    tfidf_matrix.shape}")
```

Gambar 4. Skrip ekstraksi fitur (Representasi TF-IDF)

### 4. Implementasi Perhitungan *Cosine similarity*

Setelah setiap judul skripsi diubah menjadi vektor angka menggunakan TF-IDF, tahap berikutnya adalah menghitung tingkat kemiripan antar judul menggunakan rumus *cosine similarity*. Gambar 5 menyajikan skrip yang digunakan dalam tahapan ini. Cara kerjanya adalah dengan membandingkan arah vektor dari judul baru dengan vektor- vektor judul yang sudah ada sebelumnya. Hasil perhitungan *cosine similarity* ini berupa angka antara 0 sampai 1, di mana nilai 1 berarti sangat mirip (sudut vektornya sama), sedangkan nilai 0 berarti tidak mirip sama sekali (sudut vektornya 90 derajat). Semakin mendekati angka 1, maka kedua judul tersebut dianggap semakin mirip isi atau temanya. Proses ini sangat membantu untuk mendeteksi apakah ada judul baru yang terlalu mirip dengan judul skripsi yang sudah pernah ada.

```

def detect_similarity(self, threshold=0.6):
    if self.tfidf_matrix is None:
        self.log("No TF-IDF matrix available")
        return []
    cosine_sim = cosine_similarity(self.tfidf_matrix, self.tfidf_matrix)

    similar_pairs = []

    for i in range(len(cosine_sim)):
        for j in range(i+1, len(cosine_sim)):
            similarity = cosine_sim[i][j]

            author1 = str(self.ids[i]).split('_')[0] if '_' in str(self.ids[i])
            else str(self.ids[i])
            author2 = str(self.ids[j]).split('_')[0] if '_' in str(self.ids[j])
            else str(self.ids[j])

            if similarity >= 0.999 and author1 == author2:
                continue

            if similarity >= threshold:
                similar_pairs.append({
                    'id1': self.ids[i],
                    'title1': self.titles[i],
                    'id2': self.ids[j],
                    'title2': self.titles[j],
                    'similarity': float(similarity)
                })
    return similar_pairs

```

Gambar 5. Skrip perhitungan cosine similarity

##### 5. Penentuan *Threshold* (Ambang Batas)

Setelah program menghitung skor kemiripan antar judul menggunakan *cosine similarity*, langkah berikutnya adalah menentukan *threshold* atau ambang batas. Ambang batas ini digunakan untuk memutuskan apakah sebuah judul skripsi baru dianggap mirip dengan judul yang sudah ada atau tidak. Biasanya, nilai *threshold* ini dipilih berdasarkan pengalaman atau percobaan, misalnya 0.5 atau 0.7. Artinya, jika skor kemiripan antara judul baru dan judul lama lebih besar atau sama dengan *threshold*, maka program akan menganggap judul tersebut terlalu mirip atau bahkan duplikat. Sebaliknya, jika nilainya di bawah *threshold*, maka judul dianggap berbeda dan aman untuk dipakai. *Threshold* ini penting agar program tidak terlalu sensitif (semua dibilang mirip) atau terlalu longgar (judul duplikat malah dianggap berbeda).

##### 6. Proses Utama Aplikasi

###### a. Input Judul

Tampilan yang ditunjukkan pada Gambar 6 adalah antarmuka dari sebuah sistem yang berfungsi untuk mendeteksi kemiripan judul skripsi. Sistem ini dirancang untuk membantu pengguna, khususnya mahasiswa atau dosen, dalam memeriksa apakah judul skripsi yang diusulkan memiliki kesamaan dengan judul-judul yang telah ada sebelumnya. Tujuan dari fitur ini adalah untuk menghindari duplikasi judul dan memastikan originalitas karya ilmiah yang akan dibuat.

## Deteksi Kemiripan Judul Skripsi

Masukkan judul skripsi untuk memeriksa kemiripan dengan judul yang sudah ada

Judul Skripsi

Masukkan judul skripsi yang ingin diperiksa...

Periksa Kemiripan

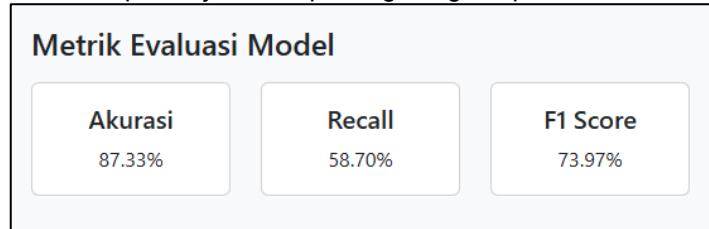
Gambar 6. Tampilan antarmuka sistem pendekripsi kemiripan judul skripsi

Pada bagian atas tampilan terdapat judul utama yang berbunyi "Deteksi Kemiripan Judul Skripsi". Judul ini memberikan informasi yang jelas mengenai fungsi dari sistem yang sedang digunakan. Tepat di bawahnya terdapat kalimat penjelas singkat yang mengarahkan pengguna untuk

memasukkan judul skripsi yang ingin diperiksa tingkat kemiripannya. Sistem ini menyediakan sebuah kolom teks dengan label "Judul Skripsi" dan petunjuk input berupa placeholder yang bertuliskan "Masukkan judul skripsi yang ingin diperiksa...". Kolom ini digunakan oleh pengguna untuk mengetikkan judul skripsi yang akan dianalisis. Setelah itu, pengguna dapat menekan tombol biru bertuliskan "Periksa Kemiripan" untuk memulai proses pemeriksaan.

b. Metrik Evaluasi Model

Gambar 7 menampilkan hasil evaluasi dari kinerja model yang digunakan untuk mendeteksi kemiripan judul skripsi. Evaluasi ini disajikan melalui tiga metrik utama, yaitu akurasi, *recall*, dan F1-score. Akurasi model sebesar 87,33% menunjukkan bahwa sebagian besar prediksi yang dilakukan oleh model sudah sesuai dengan data sebenarnya. Ini berarti model cukup handal dalam mengklasifikasikan apakah judul skripsi tergolong mirip atau tidak.



Gambar 7. Hasil evaluasi dari kinerja model

Namun, nilai *recall* yang hanya sebesar 58,70% mengindikasikan bahwa model masih melewatkannya beberapa judul yang seharusnya terdeteksi sebagai mirip. *Recall* sendiri menggambarkan kemampuan model dalam menemukan semua data yang relevan, sehingga semakin tinggi nilai *recall*, semakin baik sensitivitas model terhadap judul-judul yang mirip. Sementara itu, nilai F1-score sebesar 73,97% menunjukkan keseimbangan antara presisi dan *recall*. Nilai ini cukup baik, yang berarti model tidak hanya cukup akurat dalam hasilnya, tetapi juga relatif seimbang dalam mendeteksi dan menghindari kesalahan.

7. Hasil Pengujian

Pengujian sistem dilakukan dengan cara menginput salah satu judul skripsi, yaitu "Pengembangan e-komik media pembelajaran menulis teks negosiasi". Judul ini dipilih untuk menguji kemampuan sistem dalam mendeteksi kemiripan dengan judul-judul skripsi lain yang telah tersimpan dalam basis data. Setelah proses input dilakukan, sistem secara otomatis mengonversi judul tersebut ke dalam representasi vektor melalui metode TF-IDF, kemudian menghitung skor kemiripan dengan seluruh judul yang ada di dalam database menggunakan pendekatan *Cosine similarity*.

### Hasil Pemeriksaan Kemiripan

Judul yang diperiksa: "pengembangan e-komik media pembelajaran menulis teks negosiasi"

Judul yang Mirip:	
PENINGKATAN PEMBELAJARAN MENULIS TEKS PROSEDUR KOMPLEKS DENGAN MENGGUNAKAN PENDEKATAN PROSES BAGI SISWA SMK NEGERI 2 BUNGORO	34.7% Mirip
ID: SUMIATI	
KEEFKTIFAN MODEL PEMBELAJARAN BERBASIS PROYEK PADA PEMBELAJARAN MENULIS TEKS NEGOSIASI SISWA KELAS X MA AISIYAH SUNGGUMINASA	34.4% Mirip
ID: NURLAELA	
EFEKTIVITAS IMPLEMENTASI MODEL PEMBELAJARAN BERBASIS MASALAH PADA PEMBELAJARAN MENULIS TEKS EKSPOSISI DEFINISI SISWA KELAS VII SMP NEGERI1 SUNGGUMINASA KABUPATEN GOWA	33.5% Mirip
ID: ASDAR	
Pengaruh Penggunaan Strategi Think-Talk-Write dalam Pembelajaran Menulis Teks Berita pada Siswa Kelas VIII SMP Muhammadiyah Bungoro	29.2% Mirip
ID: JABAL NUR	

Gambar 8. Hasil deteksi kemiripan judul

Sebagai hasil dari proses tersebut, sistem menghasilkan daftar judul skripsi yang dianggap memiliki tingkat kemiripan tertentu dengan judul yang diuji. Lima judul dengan skor kemiripan tertinggi ditampilkan kepada pengguna, lengkap dengan persentase kemiripannya seperti yang

ditampilkan pada Gambar 8. Sebagai contoh, judul dengan skor tertinggi adalah “Peningkatan Pembelajaran Menulis Teks Prosedur Kompleks dengan Menggunakan Pendekatan Proses bagi Siswa SMK Negeri 2 Bungoro” yang memperoleh skor kemiripan sebesar 34.7%. Judul lainnya seperti “Keefektifan Model Pembelajaran Berbasis Proyek pada Pembelajaran Menulis Teks Negosiasi” dan “Efektivitas Implementasi Model Pembelajaran Berbasis Masalah pada Teks Eksposisi” juga terdeteksi dengan skor mendekati 30%. Hal ini membuktikan bahwa sistem mampu menangkap hubungan semantik antarjudul meskipun tidak identik secara leksikal. Selanjutnya, untuk mengevaluasi performa model secara menyeluruh, digunakan empat metrik utama, yaitu akurasi, presisi, *recall*, dan F1-score. Gambar 9 menampilkan hasil evaluasi, dan diperoleh akurasi sebesar 87.33%, yang menunjukkan bahwa sebagian besar prediksi sistem telah sesuai dengan data yang sebenarnya. Nilai presisi mencapai 100.00%, menandakan bahwa semua prediksi kemiripan yang dibuat oleh sistem benar-benar tepat tanpa adanya kesalahan klasifikasi positif palsu (false positive). Namun demikian, *recall* sistem masih berada pada angka 58.70%, yang berarti masih terdapat sejumlah judul mirip yang belum berhasil terdeteksi. Untuk mengukur keseimbangan antara presisi dan *recall*, digunakan metrik F1-Score yang berada pada nilai 73.97% (lihat Gambar 7), menunjukkan performa model berada pada tingkat yang cukup baik.

Metrik Evaluasi Model			
Akurasi	Presisi	Recall	F1 Score
87.33%	100.00%	58.70%	73.97%

Gambar 9. Hasil evaluasi dari kinerja model

Secara keseluruhan, hasil pengujian ini menunjukkan bahwa sistem deteksi kemiripan judul skripsi berbasis pendekatan semantik TF-IDF dan *Cosine similarity* telah mampu berfungsi dengan baik, khususnya dalam hal presisi. Namun, perbaikan pada aspek *recall* masih diperlukan agar sistem dapat mengenali lebih banyak judul yang memiliki kemiripan semantik secara menyeluruh. Dengan demikian, sistem ini sangat potensial untuk digunakan sebagai alat bantu dalam validasi keaslian judul skripsi dan pencegahan duplikasi topik di lingkungan akademik.

### 3.2. Pembahasan

Gambar 9 menunjukkan hasil akhir dari implementasi sistem deteksi kemiripan judul skripsi yang telah dikembangkan dalam penelitian ini. Sistem ini dirancang untuk membantu pengguna, khususnya mahasiswa dan dosen pembimbing, dalam memverifikasi tingkat kemiripan judul skripsi baru terhadap database skripsi yang telah ada. Fitur ini berfungsi sebagai alat bantu dalam mencegah pengajuan judul yang bersifat duplikat atau terlalu mirip, sehingga dapat mendorong terciptanya orisinalitas dalam penulisan karya ilmiah di lingkungan akademik.

Melalui antarmuka sistem, pengguna dapat memasukkan judul skripsi yang ingin diperiksa. Setelah dilakukan proses pemeriksaan, sistem akan menampilkan daftar judul-judul lain yang dianggap memiliki kemiripan, lengkap dengan persentase tingkat kemiripannya.

Dalam contoh hasil pengujian yang ditampilkan, beberapa judul memiliki tingkat kemiripan mulai dari 22% hingga 34%. Informasi ini dilengkapi dengan nama penulis dari masing-masing judul skripsi yang ditemukan. Nilai persentase tersebut dihitung menggunakan metode *cosine similarity* terhadap representasi vektor teks hasil dari algoritma TF-IDF. Semakin tinggi nilai persentase kemiripan, maka semakin besar kemungkinan kedua judul memiliki kesamaan dalam hal kata kunci, struktur kalimat, atau bahkan topik bahasan.

Di bagian bawah tampilan sistem, disajikan pula hasil evaluasi performa dari model klasifikasi yang digunakan. Evaluasi dilakukan dengan menggunakan empat metrik utama, yaitu akurasi, presisi, *recall*, dan F1-score. Hasil evaluasi menunjukkan bahwa akurasi model mencapai 87,33%, menandakan bahwa sebagian besar prediksi sistem sudah sesuai dengan kategori sebenarnya, baik saat mengidentifikasi judul yang mirip maupun tidak. Nilai presisi mencapai 100%, yang berarti semua judul yang diklasifikasikan sebagai mirip benar-benar relevan tanpa adanya kesalahan klasifikasi positif (false positive). Hal ini menunjukkan keandalan sistem dalam memberikan rekomendasi hasil yang benar-benar mirip.

Namun, *recall* sistem yang berada pada angka 58,70% mengindikasikan masih adanya beberapa judul yang sebenarnya mirip, tetapi tidak berhasil dikenali oleh sistem (false negative).

Rendahnya nilai *recall* ini kemungkinan disebabkan oleh keterbatasan representasi semantik pada metode TF-IDF, yang cenderung berfokus pada frekuensi kata dan belum sepenuhnya mampu menangkap konteks kalimat secara utuh. Variasi dalam gaya bahasa, penggunaan sinonim, atau struktur kalimat yang kompleks juga menjadi tantangan dalam proses deteksi kemiripan berbasis kata. Oleh karena itu, meskipun sistem sudah sangat baik dalam menghindari kesalahan positif, masih terdapat ruang untuk meningkatkan sensitivitas sistem dalam mendeteksi seluruh kemiripan yang relevan.

Untuk mengevaluasi keseimbangan antara presisi dan *recall*, digunakan metrik F1-score, yang dalam sistem ini tercatat sebesar 73,97%. Nilai ini mencerminkan bahwa sistem telah mencapai performa yang cukup baik secara keseluruhan. Akan tetapi, peningkatan pada aspek *recall* tetap menjadi fokus penting untuk pengembangan sistem di masa mendatang, agar sistem tidak hanya tepat dalam hasil yang diberikan, tetapi juga lengkap dalam cakupan deteksi kemiripan judul

#### 4. Kesimpulan

Penelitian ini berhasil mencapai tujuan yang telah ditetapkan dalam bagian pendahuluan, yaitu mengimplementasikan metode *cosine similarity* berbasis representasi TF-IDF untuk mendeteksi kemiripan judul skripsi mahasiswa FKIP Unismuh Makassar serta mengevaluasi kinerjanya secara komprehensif. Sistem yang dikembangkan mampu menerima input judul skripsi baru dan membandingkannya dengan database judul yang telah ada, kemudian menampilkan hasil berupa persentase kemiripan yang mudah dipahami oleh pengguna. Hasil evaluasi kinerja menunjukkan bahwa metode *cosine similarity* memiliki kelebihan yang signifikan dalam hal presisi, yang mencapai nilai sempurna sebesar 100%, menandakan bahwa sistem sangat andal dalam mengidentifikasi judul-judul yang benar-benar mirip tanpa menghasilkan kesalahan positif palsu. Akurasi sistem yang mencapai 87,33% juga menunjukkan bahwa sebagian besar prediksi telah sesuai dengan kondisi sebenarnya, sedangkan nilai F1-score sebesar 73,97% mencerminkan keseimbangan yang cukup baik antara presisi dan *recall*. Namun demikian, penelitian ini juga mengidentifikasi keterbatasan utama metode *cosine similarity*, yaitu nilai *recall* yang hanya mencapai 58,70%, yang mengindikasikan bahwa sistem belum sepenuhnya mampu mendeteksi semua judul yang sebenarnya memiliki kemiripan semantik. Keterbatasan ini terutama disebabkan oleh karakteristik TF-IDF yang berfokus pada pencocokan kata secara leksikal dan belum mampu menangkap kesamaan makna yang diungkapkan dengan kata-kata berbeda atau penggunaan sinonim, sebagaimana telah diidentifikasi dalam penelitian terdahulu. Berdasarkan temuan ini, disarankan agar pengembangan sistem di masa mendatang mempertimbangkan penggunaan pendekatan berbasis word embedding seperti Word2Vec atau model transformer seperti BERT yang dapat menangkap kesamaan semantik secara lebih baik, serta melakukan eksperimen dengan berbagai nilai *threshold* untuk menemukan keseimbangan optimal antara sensitivitas dan spesifisitas deteksi. Selain itu, perlu dilakukan pengujian dengan dataset yang lebih besar dan beragam serta penambahan fitur *preprocessing* yang lebih canggih seperti pengenalan frasa kunci dan penanganan variasi penulisan istilah akademik untuk meningkatkan kemampuan sistem dalam mendeteksi kemiripan judul secara menyeluruh.

#### Referensi

- [1] M. A. Shadiqi, "Memahami dan Mencegah Perilaku Plagiarisme dalam Menulis Karya Ilmiah," *Bul. Psikol.*, vol. 27, no. 1, hal. 30, 2019, doi: 10.22146/buletinpsikologi.43058.
- [2] F. Syah, N. Saputra, dan S. Maharani, "Sistem Deteksi Kemiripan Judul Skripsi dengan Metode Term Weight dan Word Embedding di Program Studi Informatika Universitas PGRI Yogyakarta," *J. Din. Inform.*, vol. 9, no. 2, hal. 1–15, 2022.
- [3] S. Ansisi, E. P. Listyaningsih, dan H. Soetanto, "Deteksi Plagiat Tesis Berbahasa Indonesia Menggunakan Metode Cosine Similarity," *INOVTEK Polbeng - Seri Inform.*, vol. 9, no. 1, hal. 153–167, 2024, doi: 10.35314/isi.v9i1.4003.
- [4] A. Arsal, M. Hamid, dan M. Santosa, "Penerapan Teks Mining Dan Cosine Similarity Untuk Menentukan Kesamaan Dokumen Skripsi Application of Text Mining and Cosine Similarity To Determine the Similarity of Thesis Documents," *IJIS Indones. J. Inf. Syst.*, vol. 9, no. April 2024, hal. 99–109, 2024.
- [5] Y. vita Via dan R. Mumpuni, "Deteksi Kemiripan Dokumen Publikasi Skripsi Mahasiswa Menggunakan Algoritma Modifikasi Cosine Similarity," *J. Inf. Eng. Educ. Technol.*, vol. 3, no. 2, hal. 57–61, Des 2019, doi: 10.26740/jieet.v3n2.p57-61.
- [6] N. Fitrianingsih, M. Asfi, D. Prasetyo, R. P. Kusuma, dan M. A. Sulhan, "Deteksi Tingkat

Kemiripan Judul Menggunakan Algoritma Oliver Pada Sistem Informasi Pengajuan Skripsi," *J. MEDIA Inform. BUDIDARMA*, vol. 6, no. 4, hal. 2148–2156, Okt 2022, doi: 10.30865/mib.v6i4.4409.

[7] N. Vendyansyah dan Y. A. Pranoto, "Perancangan dan Pembuatan Aplikasi untuk Mendeteksi Kemiripan Jawaban Menggunakan Cosine Similarity," *J. Tek. (Jurnal Fak. Tek. Univ. Islam Lamongan)*, vol. 13, no. 1, hal. 23–28, 2021.

[8] H. Herlambang, J. Suwita, dan B. Tiara, "Analisa dan Perancangan Sistem Pendekripsi Plagiarisme Skripsi pada STMIK Insan Pembangunan Menggunakan Metode Cosine Similarity," *Insa. Pembang. Sist. Inf. dan Komput.*, vol. 9, no. 1, Jun 2021, doi: 10.58217/ipsikom.v9i1.188.

[9] A. Riyani, M. Zidny Naf'an, dan A. Burhanuddin, "Penerapan Cosine Similarity dan Pembobotan TF-IDF untuk Mendeteksi Kemiripan Dokumen," *J. Linguist. Komputasi*, vol. 2, no. 1, hal. 23–27, 2019.

[10] T. A. Prismadana, "Aplikasi Ruang Tugas Dengan Deteksi Kemiripan Teks Pada Dokumen Tugas Menggunakan Cosine Similarity," *J. Inform. dan Multimed.*, vol. 15, no. 1, hal. 31–37, 2023, doi: 10.33795/jtim.v15i1.4405.

[11] H. E. U. A. Basuki, Kusrini, dan A. Nasiri, "Studi dan Analisis Algoritma Kemiripan Cosine Similarity Untuk Mengidentifikasi Kesamaan Judul Skripsi," in *Prosiding Seminar Ilmiah Sistem Informasi Dan Teknologi Informasi*, 2023, hal. 1–9.

[12] S. H. Sihombing, M. Ayo, dan E. Nasution, "Implementasi Metode Cosine Similarity Untuk Deteksi Kemiripan Data Pada Dokumen Word Dan Perbedaan Isi Dokumen Word Implementation Of The Cosine Similarity Method For Detecting Data Similarity In Word Documents And Differences In The Contents Of Word Docu," *J. Info Digit*, vol. 2, no. 2, hal. 736–748, 2024, [Daring]. Tersedia pada: <http://kti.potensi-utama.ac.id/index.php/JID>

[13] D. Kurniadi, S. F. C. Haviana, dan A. Novianto, "Implementasi Algoritma Cosine Similarity pada sistem arsip dokumen di Universitas Islam Sultan Agung," *J. Transform.*, vol. 17, no. 2, hal. 124, Jan 2020, doi: 10.26623/transformatika.v17i2.1613.

[14] H. S. Sutikno, "Implementasi Algoritma Cosine Similarity Untuk Mendeteksi Kemiripan Topik Judul," *J. Eng. Comput. Sci. Inf. Technol.*, vol. 1, no. 1, hal. 51–61, Apr 2021, doi: 10.33365/jecsit.v1i1.6.

[15] K. Telaumbanua dan L. Nababan, "Implementasi Metode Cosine Similarity Dalam Mendeteksi Kemiripan Dan Perbedaan Gambar Hasil Scan Berbasis Android," *Informatics Eng. Electron. Data*, vol. 1, no. 1, hal. 27–36, 2022, doi: 10.59840/ieed.v1i1.176.