JUSIFOR: Jurnal Sistem Informasi dan Informatika
Vol. 4, No. 2, Desember 2025, hlmn. 243-252
E-ISSN: 2830-3393 | P-ISSN: 2830-2443
DOI: https://doi.org/10.70609/jusifor.v4i2.8340

Sistem Rekomendasi Referensi Jurnal Untuk Penentuan Judul Skripsi Menggunakan Cosine Similarity Pada Data Jurnal JATI ITN Malang

Marita Putri Nabila[1*], Joseph Dedy Irawan[2], Ahmad Faisol[3]

[1,2,3] Teknik Informatika, Fakultas Teknologi industri, Institut Teknologi Nasional Malang, Indonesia

Info Artikel

Riwayat Artikel:
Diterima : **1 November 2025**
Direvisi : **19 Desember 2025**
Diterbitkan : **31 Desember 2025**

Kata Kunci:

Sistem Rekomendasi,
Cosine Similarity,
TF - IDF,
Referensi Jurnal,
Judul Skripsi

ABSTRAK

Pertumbuhan publikasi ilmiah yang meningkat di lingkungan kampus mendorong mahasiswa untuk menghasilkan skripsi yang orisinal dan relevan. Namun, meningkatnya jumlah publikasi sering membuat mahasiswa kesulitan menentukan judul skripsi karena pencarian referensi masih sederhana. Penelitian ini bertujuan untuk membangun sistem rekomendasi judul skripsi dengan algoritma *Cosine similarity* untuk membantu mahasiswa memperoleh referensi dalam menentukan judul skripsi serta mencegah duplikasi topik penelitian. Sistem ini menggunakan dataset publikasi JATI (Jurnal Mahasiswa Teknik Informatika) ITN Malang, melalui tahapan preprocessing teks, pembobotan TF-IDF, serta perhitungan *cosine similarity* dalam mengukur tingkat kemiripan judul publikasi di JATI ITN Malang. Evaluasi dilakukan dengan metode *Confusion Matrix* untuk menilai performa sistem dalam memberikan rekomendasi yang relevan. Hasil pengujian menunjukkan nilai akurasi sebesar 88,89%, *precision* 100%, *recall* 83,33%, dan *F1-score* 90,91%. Berdasarkan hasil yang diperoleh, sistem terbukti mampu merekomendasikan judul skripsi yang sesuai, sehingga dapat membantu mahasiswa dalam proses penentuan judul sebagai acuan penyusunan skripsi.

***Keywords*:**

Recommendation System,
Cosine Similarity,
TF - IDF,
Journal Reference,
Thesis Title

ABSTRACT

The increasing growth of scientific publications within university environments encourages students to produce original and relevant undergraduate theses. However, the rising number of publications often makes it difficult for students to determine thesis titles, as reference searches are still relatively simple. This study aims to develop a thesis title recommendation system using the Cosine similarity algorithm to assist students in obtaining references for determining their thesis topics and to prevent topic duplication. The system utilizes a dataset from JATI (Jurnal Mahasiswa Teknik Informatika), ITN Malang, and involves several stages: text preprocessing, TF-IDF weighting, and cosine similarity computation to measure the degree of similarity among publication titles in JATI ITN Malang. The evaluation was conducted using the Confusion Matrix method to assess the system's performance in generating relevant recommendations. The experimental results show an accuracy of 88,89%, precision of 100%, recall of 83.33%, and an F1-score of 90.91%. Based on these findings, the system has proven capable of recommending appropriate thesis titles, thereby assisting students in the process of selecting and formulating their undergraduate thesis topics.

Penulis Korespondensi:
Marita Putri Nabila,
Program Studi Teknik Informatika,
Institut Teknologi Nasional Malang
Email: maritap225@gmail.com

This is an open access article under the CC BY-SA license

1. PENDAHULUAN

Pertumbuhan publikasi ilmiah di Indonesia meningkat setiap tahun, termasuk di lingkungan kampus. Salah satu sarana publikasi online adalah JATI (Jurnal Mahasiswa Teknik Informatika) ITN Malang, yang menyajikan temuan penelitian terkait teknologi informasi dan komputasi. Dalam dunia pendidikan tinggi, skripsi merupakan salah satu bentuk tugas akhir yang penting diselesaikan oleh mahasiswa dalam mendapatkan gelar sarjana [1]. Sebagai institusi pendidikan, perguruan tinggi berperan aktif dalam meningkatkan kesadaran serta memberikan panduan untuk mencegah plagiarisme publikasi [2]. Namun, dengan meningkatnya jumlah publikasi, mahasiswa sering mengalami kesulitan dalam menentukan judul skripsi yang relevan, karena pencarian referensi masih manual atau berbasis kata kunci sederhana. Fenomena tersebut menyebabkan banyak informasi yang masuk atau bisa disebut dengan *information overload* [3]. Masalah utama yang dihadapi mahasiswa adalah kesulitan dalam menentukan judul skripsi yang relevan akibat meningkatnya jumlah publikasi di JATI. Sistem pencarian sederhana yang belum mampu menilai kemiripan antar judul penelitian secara efektif. Akibatnya, mahasiswa harus menelusuri seluruh publikasi secara manual, yang memakan waktu lama dan kurang efisien.

Beberapa penelitian sebelumnya telah menggunakan metode *cosine similarity* untuk membangun sistem rekomendasi di berbagai domain. Fisabilillah dkk. mengembangkan sistem rekomendasi konversi mata kuliah MSIB di ITN Malang, yang membantu mahasiswa memilih mata kuliah berdasarkan deskripsi silabus dengan hasil akurasi 85% [4]. Vendyansyah & Pranoto merancang aplikasi untuk mendeteksi kemiripan jawaban siswa dengan kunci jawaban menggunakan *cosine similarity*, menghasilkan selisih error hanya 0,27% dibanding perhitungan manual [5]. Velamentosa & Zuliarso membangun sistem rekomendasi film berbasis CBF, yang memberikan rekomendasi spesifik untuk user [6]. Muria dkk. dengan sistem rekomendasi jurnal TIK dalam membantu memilih publikasi di laman Garuda [7]. Elnursa dkk. merancang sistem rekomendasi program MSIB mahasiswa, menghasilkan daftar rekomendasi yang relevan sesuai preferensi pengguna dengan hasil *precision* rata-rata 89,4% [8]. Namun, semua penelitian tersebut belum diterapkan untuk membantu mahasiswa menentukan judul skripsi di JATI ITN Malang secara otomatis, sehingga terdapat gap penelitian yang dapat diisi oleh penelitian ini.

Penelitian ini mengusulkan pembangunan sistem rekomendasi judul skripsi berbasis dengan penggunaan *cosine similarity* dalam menghitung skor kemiripan atau *similarity* antar judul dan kata kunci. Secara umum, sebuah sistem rekomendasi terdiri dari tiga elemen pokok, yaitu proses pengumpulan informasi, algoritma rekomendasi, dan tampilan interaksi bagi pengguna. [9]. Sistem bertujuan untuk membantu mahasiswa menghindari plagiasi atau duplikat judul, dan meningkatkan literatur referensi sebagai acuan dalam penyusunan skripsi.

Penggunaan algoritma *cosine similarity* memungkinkan sistem untuk secara objektif dan konsisten mengukur kesamaan antara dua teks [10]. Algoritma ini dipilih karena kesederhanaan dan efisiensinya dalam menghitung kesamaan teks berbasis distribusi kata [11]. Metode ini mampu mengukur tingkat kemiripan antar teks secara efektif, independen terhadap panjang dokumen, dan telah terbukti memberikan akurasi tinggi dalam sistem rekomendasi berbasis konten. Dengan metode *cosine similarity*, sistem memproses data judul dan kata kunci publikasi di JATI ITN Malang. Proses dimulai dengan preprocessing teks. Selanjutnya, teks dikonversi menjadi vektor menggunakan representasi TF-IDF, dan dilakukan perhitungan tingkat kesamaan antar dokumen berdasarkan algoritma *cosine similarity*. Dataset yang digunakan terdiri dari seluruh publikasi di JATI yang relevan. Sistem ini menghasilkan skor kesamaan untuk setiap pasangan dokumen dan merekomendasikan judul skripsi dengan skor tertinggi, sehingga mahasiswa memperoleh judul yang relevan dan memiliki literatur yang memadai.

2. METODE PENELITIAN

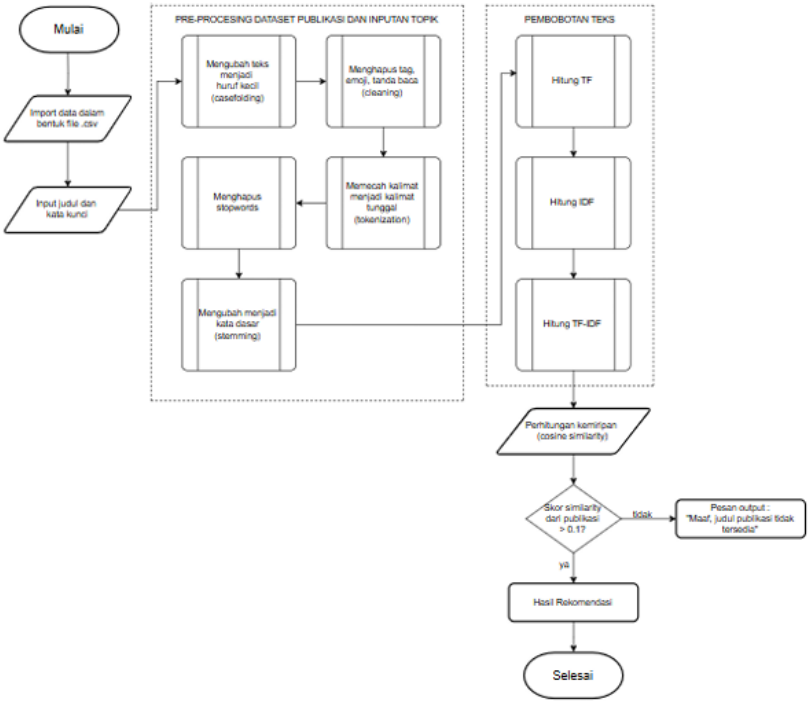
Metode penelitian ini dirancang untuk mengatasi kesulitan mahasiswa dalam menentukan judul skripsi yang relevan dengan sumber referensi publikasi di JATI ITN Malang. Sistem rekomendasi dikembangkan menggunakan pendekatan content-based filtering dengan algoritma *cosine similarity* untuk menganalisis tingkat kemiripan antar judul publikasi.

2.1 Pengumpulan Data

Data diperoleh melalui pengambilan informasi dari situs JATI ITN Malang. Kegiatan mengambil data dari sebuah website seperti ini dikenal sebagai web scraping [12]. Pengambilan data berupa publikasi ilmiah yang tersedia meliputi judul, penulis, tahun, abstrak, kata kunci, afiliasi, dan url jurnal. Pengumpulan data publikasi ini diambil mulai dari tahun 2017 hingga 2025 terbaru.

2.2 Pengolahan Data

Alur pengolahan data untuk sistem rekomendasi referensi jurnal untuk penentuan judul skripsi digambarkan pada Gambar 2 berikut. Tampilan *flowchart* ini untuk memberikan gambaran umum mengenai proses kerja sistem dari tahap awal hingga menghasilkan rekomendasi judul jurnal yang relevan. Setiap tahap dalam alur ini menunjukkan bagaimana data dikumpulkan, diproses, dan dianalisis.



Gambar 1. *Flowchart* Sistem

Berikut penjelasan untuk tahapan memperoleh rekomendasi judul skripsi :

a. Input Data

Tahap pertama yang dilakukan adalah user menginputkan data yang dibutuhkan. Untuk dataset berupa data publikasi yang telah di scraping. Untuk data inputan berupa judul dan kata kunci.

b. Pre-processing

Tahap kedua adalah pre-processing, yaitu tahapan dalam text mining untuk mendapatkan informasi dari teks dengan kualitas lebih tinggi. Preprocessing berfungsi menyiapkan teks agar dapat diolah dan dapat dianalisis lebih lanjut. Karena data berupa teks mentah tidak langsung diproses, sehingga diperlukan tahapan ini untuk mengubahnya menjadi bentuk numerik [13]. Langkah-langkah pre-processing yang dilakukan meliputi *cleaning* (mengubah data mentah ke format siap analisis [4]), *case folding* (menormalkan huruf menjadi kecil), *tokenizing* (teks diurai berbentuk kata-kata), *stopword removal* (pembersihan kata kurang bermakna), dan *stemming* (mengubah kata menjadi bentuk dasar agar teks lebih bersih dan siap dianalisis).

c. Pembobotan TF-IDF

Tahap ketiga adalah pembobotan TF-IDF, setelah data diproses selanjutnya direpresentasikan menjadi bentuk vektor menggunakan TF-IDF (*Term Frequency-Inverse Document Frequency*). Metode ini digunakan untuk mengukur bobot kata berdasarkan kepentingannya dalam suatu dokumen. Frekuensi suatu istilah dalam data ditentukan menggunakan model frekuensi kata dan frekuensi dokumen terbalik (IDF), di mana TF

mengukur berapa kali sebuah istilah muncul [14]. Semakin sering sebuah term muncul maka akan mempengaruhi nilai pembobotan [15].

d. *Cosine Similarity*

Tahap keempat untuk menghitung tingkat kemiripan antar judul dan kata kunci, digunakan algoritma *cosine similarity*. Algoritma ini untuk menghtung kesamaan antar dua vektor berdasarkan sudutnya. Semakin kecil sudut, semakin mirip keduanya. Ini sering digunakan untuk mengukur kesamaan makna antar objek [4]. Berikut rumus *cosine similarity* :

$$Similarity = cos(\theta) = \frac{A \cdot B}{\|A\|\|B\|} = \frac{\left(\sum_{i=1}^{n} A_i \cdot B_i\right)}{\left(\sqrt{\sum_{i=1}^{n} (A_i)^2}\sqrt{\sum_{i=1}^{n} (B_i)^2}\right)} \quad (1)$$

Keterangan :

A : vektor representasi judul skripsi
B : vektor representasi kata kunci jurnal
A_i: bobot kata ke-i pada vektor judul
B_i: bobot kata ke-i pada vektor kata kunci
n : jumlah total kata (fitur) yang digunakan
A·B: perkalian titik (dot product) antara vektor A dan B
‖A‖: panjang (norma) vektor A, dihitung dari akar jumlah kuadrat setiap elemennya
‖B‖: panjang (norma) vektor B, dihitung dari akar jumlah kuadrat setiap elemennya
cos(θ): nilai kemiripan antara dua vektor berdasarkan sudut di antaranya.

Nilai Similarity : berada pada rentang 0 – 1, semakin mendekati 1 maka semakin tinggi tingkat kemiripan

e. Rekomendasi Output

Skor kesamaan hasil perhitungan *Cosine similarity* digunakan dalam menghasilkan output rekomendasi referensi jurnal untuk judul skripsi yang relevan dengan inputan pengguna. Hasil perhitungan ini disusun secara berurutan berdasarkan nilai kesamaan tertinggi. Judul jurnal skripsi dengan tingkat kesamaan paling tinggi ditempatkan pada posisi teratas sebagai rekomendasi utama bagi pengguna.

2.3 Evaluasi

Sistem rekomendasi yang dibangun akan dilakukan evaluasi untuk mengukur kualitas rekomendasi, menggunakan *Confusion Matrix*. *Confusion Matrix* merupakan bagan untuk menampilkan total dokumen uji yang dikategorikan benar dan salah, serta meringkas hasil prediksi metode pengelompokan[14]. Matrik yang digunakan dalam evaluasi adalah akurasi, *precision*, *recall*, dan *F1-score* [4].

a. Presisi, mengukur ketepatan hasil rekomendasi yang relevan.

b. *Recall*, menilai kemampuan sistem menemukan seluruh data relevan.

c. *F1-score*, keseimbangan antara presisi dan *recall*.

d. Akurasi, menunjukkan tingkat kebenaran keseluruhan rekomendasi yang dihasilkan sistem.

3. HASIL DAN ANALISIS

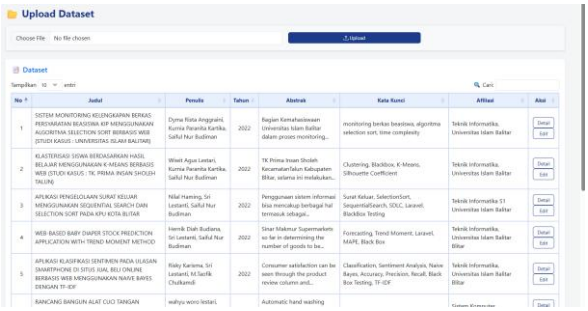
3.1 Hasil Pengumpulan Data

Berdasarkan data yang diperoleh melalui scraping website e-journal JATI ITN Malang, didapatkan data publikasi sebanyak 4091 dokumen. Setiap dokumen mencakup data yang meliputi judul, penulis, tahun, abstrak, kata kunci, afiliasi, dan url jurnal. Data disimpan dalam format CSV agar memudahkan dalam pemrosesan lebih lanjut.

3.2 Input Dataset

Tampilan input dataset dilakukan oleh admin sistem yang ditunjukkan pada Gambar 2 berikut. Terdiri dari tombol pilih file dengan format CSV lalu tombol upload untuk menampilkan dataset dalam bentuk tabel di

bawahnya. Dataset ini yang digunakan dalam tahapan proses untuk menghasilkan rekomendasi referensi jurnal untuk judul skripsi, khususnya data judul dan kata kunci.

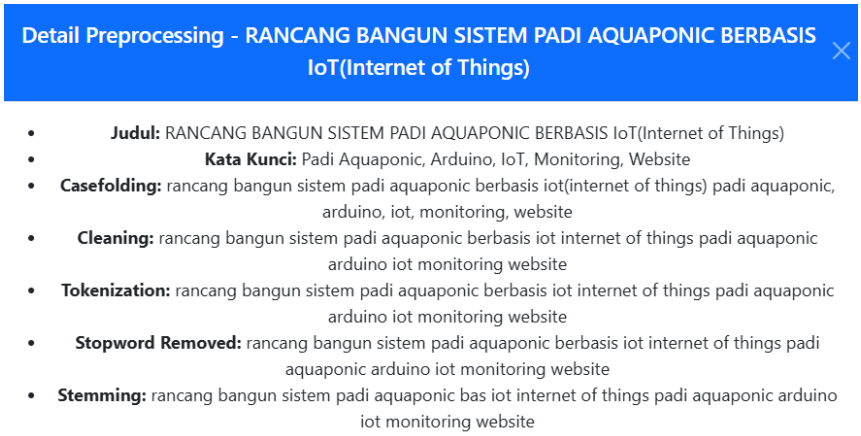


Gambar 2. Tampilan Input Dataset

Setelah file diunggah, data akan ditampilkan dalam bentuk tabel yang memuat informasi seperti nomor, judul, penulis, tahun, abstrak, kata kunci, dan afiliasi. Setiap baris data juga memiliki tombol Detail untuk melihat informasi lebih lengkap tentang data jurnal tersebut. Data ini akan disimpan di database MySQL untuk proses berikutnya.

3.3 Pre-processing

Hasil pre-processing berikut adalah teks yang melalui beberapa tahapan yang telah dibersihkan dari karakter khusus, tanda baca, angka, stopwords, dan diubah menjadi huruf kecil. Hasil teks berupa gabungan processed text dari judul dan kata kunci yang sudah dilakukan tahapan pre-processing. Contoh hasil preprocessing dari satu dokumen akan ditunjukkan sebagai berikut.

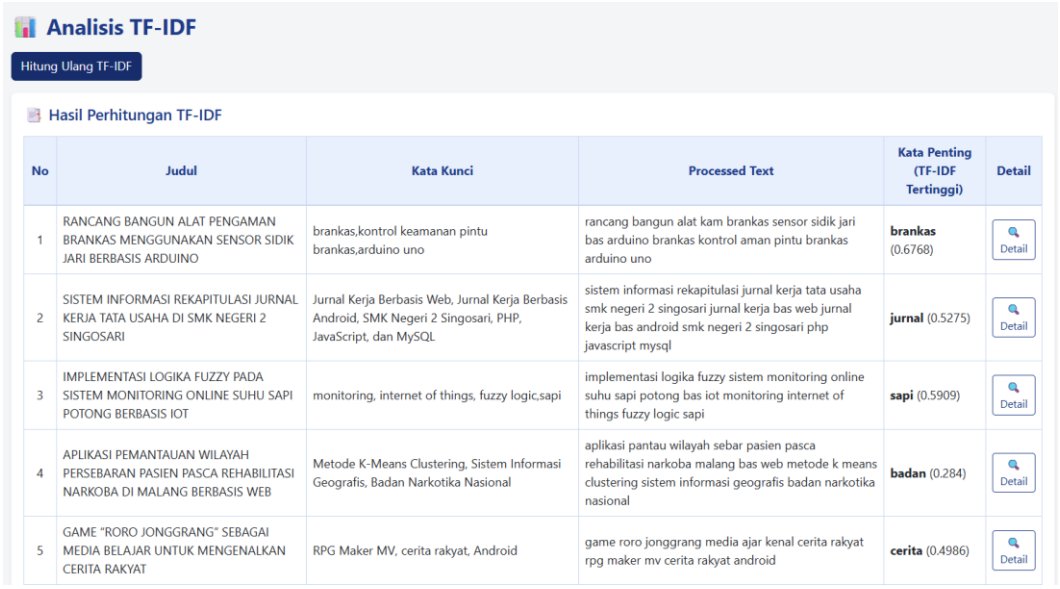


Gambar 3. Tampilan Hasil Preprocessing Teks

Halaman ini berfungsi untuk melakukan proses pembersihan teks atau pengolahan awal terhadap data judul yang telah diunggah sebelumnya. Tombol Detail disediakan di setiap baris judul jurnal agar admin dapat melihat hasil pemrosesan teks secara lebih lengkap. Fitur ini membantu memastikan bahwa data siap digunakan untuk tahap analisis berikutnya seperti perhitungan TF-IDF dan cosine similarity.

3.4 Pembobotan TF-IDF

Hasil pembobotan kata dengan metode TF-IDF dalam menentukan tingkat kepentingan kata pada dokumen ditunjukkan pada gambar 4 berikut. Nilai TF-IDF yang tinggi menandakan kata lebih berpengaruh dalam merepresentasikan isi dokumen. Nilai TF-IDF sebesar 0.6270 menunjukkan bahwa kata "Aquaponic" memiliki bobot yang cukup tinggi pada dokumen tersebut karena sering muncul dalam dokumen (TF cukup besar), namun jarang muncul pada dokumen lain (DF kecil → IDF besar).



Analisis TF-IDF

Hitung Ulang TF-IDF

Hasil Perhitungan TF-IDF

No	Judul	Kata Kunci	Processed Text	Kata Penting (TF-IDF Tertinggi)	Detail
1	RANCANG BANGUN ALAT PENGAMAN BRANKAS MENGGUNAKAN SENSOR SIDIK JARI BERBASIS ARDUINO	brankas,kontrol keamanan pintu brankas,arduino uno	rancang bangun alat kam brankas sensor sidik jari bas arduino brankas kontrol aman pintu brankas arduino uno	**brankas** (0.6768)	Detail
2	SISTEM INFORMASI REKAPITULASI JURNAL KERJA TATA USAHA DI SMK NEGERI 2 SINGOSARI	Jurnal Kerja Berbasis Web, Jurnal Kerja Berbasis Android, SMK Negeri 2 Singosari, PHP, JavaScript, dan MySQL	sistem informasi rekapitulasi jurnal kerja tata usaha smk negeri 2 singosari jurnal kerja bas web jurnal kerja bas android smk negeri 2 singosari php javascript mysql	**jurnal** (0.5275)	Detail
3	IMPLEMENTASI LOGIKA FUZZY PADA SISTEM MONITORING ONLINE SUHU SAPI POTONG BERBASIS IOT	monitoring, internet of things, fuzzy logic,sapi	implementasi logika fuzzy sistem monitoring online suhu sapi potong bas iot monitoring internet of things fuzzy logic sapi	**sapi** (0.5909)	Detail
4	APLIKASI PEMANTAUAN WILAYAH PERSEBARAN PASIEN PASCA REHABILITASI NARKOBA DI MALANG BERBASIS WEB	Metode K-Means Clustering, Sistem Informasi Geografis, Badan Narkotika Nasional	aplikasi pantau wilayah sebar pasien pasca rehabilitasi narkoba malang bas web metode k means clustering sistem informasi geografis badan narkotika nasional	**badan** (0.284)	Detail
5	GAME "RORO JONGGRANG" SEBAGAI MEDIA BELAJAR UNTUK MENGENALKAN CERITA RAKYAT	RPG Maker MV, cerita rakyat, Android	game roro jonggrang media ajar kenal cerita rakyat rpg maker mv cerita rakyat android	**cerita** (0.4986)	Detail

Gambar 4. Tampilan Hasil TF-IDF

Hasil pembobotan ini menjadi dasar dalam tahap perhitungan kesamaan dokumen menggunakan metode Cosine similarity. Berikut contoh perhitungan pembobotan TF-IDF dengan jumlah data sebanyak 110 data.

Tabel 2. Perhitungan TF-IDF

Kata	Count	DF	TF	IDF	TF-IDF
Aquaponic	2	1	2/16 = 0.1250	Idf = ln (1+110 / 1+1) + 1 = ln (111 / 2) + 1 = 5.016383	0.1250 x 5.016383 = 0.6270

Tabel 3. Hasil Pembobotan TF-IDF

Kata	Nilai TF-IDF	Kata	Nilai TF-IDF	Kata	Nilai TF-IDF
Aquaponic	0.6270	Arduino	0.1681	Sistem	0.1089
Bangun	0.1651	Bas	0.1179	Website	0.1713
Internet	0.1568	Iot	0.3037	Things	0.1651
Monitoring	0.1543	Of	0.1371		
Padi	0.5357	Rancang	0.1713		

3.5 Cosine Similarity

a. User menginputkan judul dan kata kunci, lalu akan melalui proses preprocessing dan pembobotan tf-idf.

b. Hasil rekomendasi akan menampilkan top 3 judul dengan similarity tertinggi.

c. Proses perhitungan cosine similarity sendiri membandingkan nilai vector TF-IDF, antara inputan user dan dokumen dataset. Dimana dalam rumus cosine similarity terdapat vektor A (vector TF-IDF input user) dan vector B (vector TF-IDF salah satu dokumen dataset).

3.6 Implementasi Sistem

a. Halaman Input User

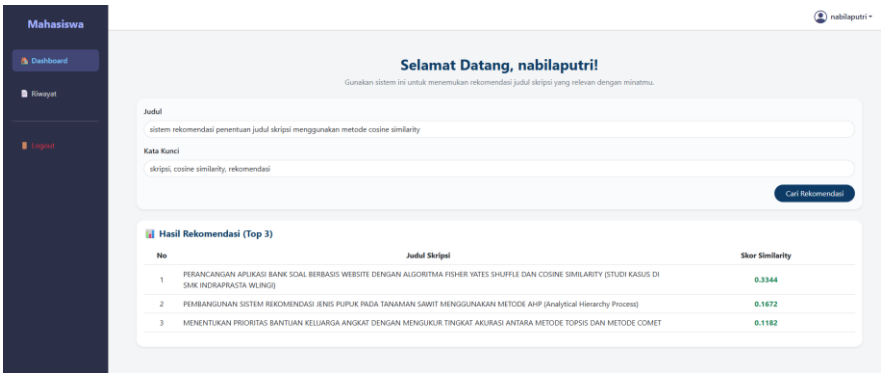
Pada Gambar 5 berikut ditunjukkan halaman utama sistem rekomendasi judul skripsi pada sisi pengguna (mahasiswa). Pada halaman ini, pengguna dapat memasukkan judul sementara (rencana judul) dan kata kunci yang sesuai dengan topik penelitian yang dicari. Setelah itu, pengguna dapat menekan tombol "Cari Rekomendasi" untuk menampilkan hasil rekomendasi judul yang relevan.

Gambar 5. Tampilan Halaman Inputan User

Tampilan ini dirancang dengan antarmuka sederhana dan mudah digunakan agar pengguna dapat dengan cepat memahami cara kerja sistem. User bisa menginputkan sesuai kebutuhan, dan saat button diklik maka akan menampilkan beberapa rekomendasi judul jurnal yang diambil dari dataset dilengkapi dengan skor similarity mulai dari yang tertinggi.

b. Tampilan Hasil Rekomendasi

Pada Gambar 6 berikut ditunjukkan hasil rekomendasi judul skripsi berdasarkan input judul dan kata kunci yang dimasukkan oleh pengguna. Sistem menampilkan tiga rekomendasi teratas (top 3) yang memiliki tingkat kesamaan / *similarity* tertinggi dengan input pengguna berdasarkan perhitungan *cosine similarity*. Setiap hasil rekomendasi disertai dengan nilai skor *similarity*, dimana nilai atau skor yang lebih tinggi menandakan tingkat relevansi yang lebih besar.

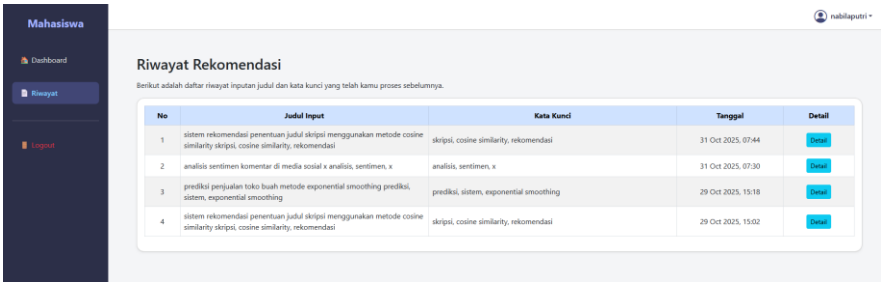


Gambar 6. Tampilan Hasil Rekomendasi

Tampilan ini membantu pengguna menentukan judul skripsi sesuai dengan topik penelitian yang dimasukkan, serta hasil rekomendasi referensi jurnal bisa dijadikan referensi isi skripsi serta menghindari plagiasi judul.

c. Halaman Riwayat Inputan User

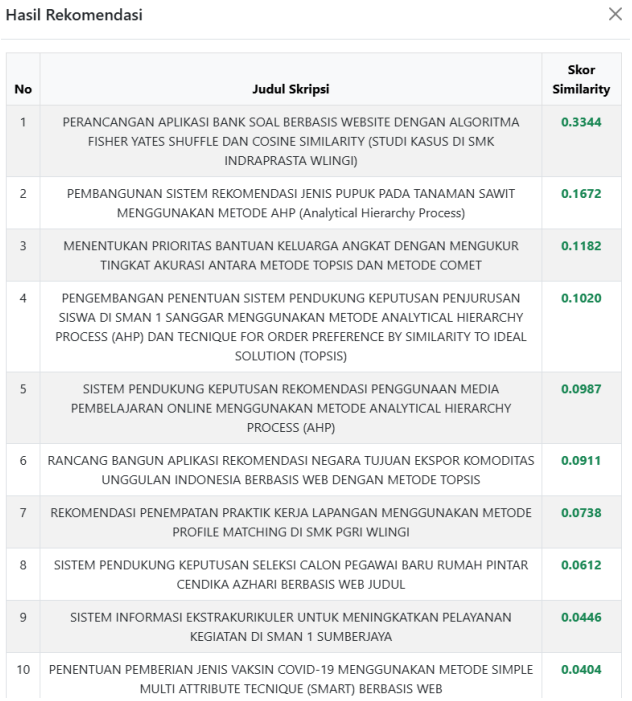
Pada Gambar 7 berikut ditunjukkan riwayat inputan pengguna, untuk menyimpan dan memudahkan jika pengguna akan mencari inputan terkait. Data inputan disimpan dan ditampilkan dalam bentuk tabel dengan kolom judul, kata kunci, dan detail. Untuk tombol detail apabila diklik akan menampilkan modal berupa hasil rekomendasi judul publikasi lengkap dengan skor *similarity* mulai dari yang tertinggi.



Gambar 7. Tampilan Riwayat Inputan User

Pada Gambar 8 berikut menampilkan hasil rekomendasi berupa list judul publikasi di JATI ITN Malang, lengkap dengan skor *similarity* mulai dari yang tertinggi hingga terendah. Output ini menghasilkan 10

rekomendasi judul jurnal skripsi dengan skor tertinggi, sesuai dengan inputan kata yang dimasukkan pengguna, sehingga menghasilkan banyak rekomendasi judul karena terdapat kata yang sama seperti inputan.



Hasil Rekomendasi

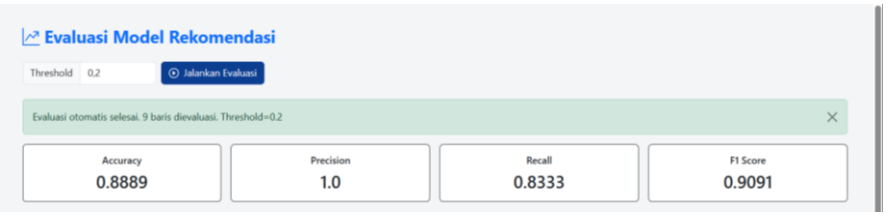
No	Judul Skripsi	Skor Similarity
1	PERANCANGAN APLIKASI BANK SOAL BERBASIS WEBSITE DENGAN ALGORITMA FISHER YATES SHUFFLE DAN COSINE SIMILARITY (STUDI KASUS DI SMK INDRAPRASTA WLINGI)	0.3344
2	PEMBANGUNAN SISTEM REKOMENDASI JENIS PUPUK PADA TANAMAN SAWIT MENGGUNAKAN METODE AHP (Analytical Hierarchy Process)	0.1672
3	MENENTUKAN PRIORITAS BANTUAN KELUARGA ANGKAT DENGAN MENGUKUR TINGKAT AKURASI ANTARA METODE TOPSIS DAN METODE COMET	0.1182
4	PENGEMBANGAN PENENTUAN SISTEM PENDUKUNG KEPUTUSAN PENJURUSAN SISWA DI SMAN 1 SANGGAR MENGGUNAKAN METODE ANALYTICAL HIERARCHY PROCESS (AHP) DAN TECNIQUE FOR ORDER PREFERENCE BY SIMILARITY TO IDEAL SOLUTION (TOPSIS)	0.1020
5	SISTEM PENDUKUNG KEPUTUSAN REKOMENDASI PENGGUNAAN MEDIA PEMBELAJARAN ONLINE MENGGUNAKAN METODE ANALYTICAL HIERARCHY PROCESS (AHP)	0.0987
6	RANCANG BANGUN APLIKASI REKOMENDASI NEGARA TUJUAN EKSPOR KOMODITAS UNGGULAN INDONESIA BERBASIS WEB DENGAN METODE TOPSIS	0.0911
7	REKOMENDASI PENEMPATAN PRAKTIK KERJA LAPANGAN MENGGUNAKAN METODE PROFILE MATCHING DI SMK PGRI WLINGI	0.0738
8	SISTEM PENDUKUNG KEPUTUSAN SELEKSI CALON PEGAWAI BARU RUMAH PINTAR CENDIKA AZHARI BERBASIS WEB JUDUL	0.0612
9	SISTEM INFORMASI EKSTRAKURIKULER UNTUK MENINGKATKAN PELAYANAN KEGIATAN DI SMAN 1 SUMBERJAYA	0.0446
10	PENENTUAN PEMBERIAN JENIS VAKSIN COVID-19 MENGGUNAKAN METODE SIMPLE MULTI ATTRIBUTE TECNIQUE (SMART) BERBASIS WEB	0.0404

Gambar 8. Tampilan Rekomendasi Judul Skripsi

Pada tiap riwayat judul jurnal yang muncul di tabel tersebut juga dilengkapi dengan link untuk akses langsung ke jurnal nya, sehingga bisa dicek tanpa mencari manual lagi. Melalui sistem rekomendasi ini, diharapkan mampu membantu mahasiswa dalam menentukan judul skripsi yang sesuai dengan topik yang diteliti, dengan rekomendasi jurnal skripsi dari JATI ITN Malang sebagai referensi untuk menentukan judul dan referensi terkait metode yang digunakan.

3.7 Evaluasi

Tahap evaluasi dilakukan untuk mengetahui sejauh mana sistem rekomendasi bekerja dengan baik dalam memberikan hasil yang relevan terhadap data uji. Evaluasi dilakukan dengan menggunakan Confusion Matrix, serta menghitung nilai Accuracy, Precision, Recall, dan F1-Score. Pada halaman ini berisi threshold yang bisa diatur, tabel hasil confusion matrix, nilai akurasi, presisi, recall, dan f1-score, serta detail hasil evaluasi. Berdasarkan hasil evaluasi, sistem memperoleh nilai akurasi 88.89%, precision 100%, recall 83.33%, dan F1-score 90.91%. Hasil tersebut menunjukkan kemampuan sistem dalam memberikan rekomendasi yang tepat dan relevan berdasarkan data yang ada.



Gambar 9. Tampilan Hasil Evaluasi

Untuk data pengujian yang digunakan adalah dataset sejumlah 110 data judul untuk dasar pelatihan model rekomendasi. Dari dataset tersebut, sistem diuji dengan 3 data input pengguna yang mewakili judul baru, dengan hasil masing-masing inputan adalah top 3 judul dataset yang dimasukkan ke dalam sistem. Pengujian berdasarkan threshold, dimana pada sistem rekomendasi ini menggunakan ambang batas (threshold) untuk menentukan kategori hasil rekomendasi. Pada pengujian ini, digunakan nilai threshold sebesar 0,2, yang berarti jika skor rekomendasi ≥ 0,2 dikategorikan sebagai positif (1) = relevan dengan topik

input pengguna, sedangkan jika skor rekomendasi < 0,2 dikategorikan sebagai negatif (0) = tidak relevan. Nilai threshold ini dipilih berdasarkan hasil eksperimen yang memberikan keseimbangan terbaik antara jumlah prediksi positif dan akurasi model.

Tabel 4. Confusion Matrix

Actual / Prediksi	Pred 0 (Tidak Relevan)	Pred 1 (Relevan)
Actual 0 (Tidak Relevan)	3	0
Actual 1 (Relevan)	1	5

Tabel 5. Perhitungan Metrik Evaluasi

Metrik	Rumus	Perhitungan	Nilai	Keterangan
Accuracy	(TP+TN)/(TP+TN+FP+ FN)	(5+3) / (5+3+0+1)	0.8889	Ketepatan keseluruhan model dalam mengklasifikasikan data.
Precision	TP/(TP+FP)	5 / (5+0)	1.0	Tingkat ketepatan sistem saat memprediksi data relevan.
Recall	TP/(TP+FN)	5 / (5+1)	0.8333	Kemampuan sistem dalam menemukan semua data relevan.
F1 – Score	$2 \times \frac{Precision \times Recall}{Precision + Recall}$	$2 \times \frac{1.0 \times 0.8333}{1.0 + 0.8333}$	0.9091	Rata-rata harmonik dari precision dan recall.

Tabel 6. Perbandingan Hasil Threshold

Threshold	Accuracy	Precision	Recall	F1-Score
0.1	0.7778	1.0	0.7778	0.875
0.2	0.8889	1.0	0.8333	0.9091
0.3	0.7778	1.0	0.6667	0.8
0.4	0.5556	1.0	0.3333	0.5
0.5	0.5556	1.0	0.3333	0.5
0.6	0.5556	1.0	0.3333	0.5
0.7	0.5556	1.0	0.3333	0.5
0.8	0.5556	1.0	0.3333	0.5
0.9	0.5556	1.0	0.3333	0.5
1.0	0.5556	1.0	0.3333	0.5

Berdasarkan hasil pengujian terhadap beberapa nilai threshold Cosine Similarity, terlihat bahwa precision tetap konstan sebesar 1.0 yang menandakan seluruh rekomendasi yang dihasilkan selalu relevan dengan kebutuhan pengguna. Namun demikian, nilai recall dan accuracy menunjukkan penurunan seiring meningkatnya threshold, sehingga sistem menjadi lebih ketat dan berpotensi mengabaikan beberapa judul yang sebenarnya relevan. Nilai F1-Score tertinggi sebesar 0.9091 diperoleh pada threshold 0.2, yang menunjukkan keseimbangan antara ketepatan dan kelengkapan rekomendasi. Dengan demikian, threshold 0.2 dinilai sebagai nilai optimal karena mampu menjaga kualitas rekomendasi tanpa mengurangi terlalu banyak alternatif yang layak. Secara keseluruhan, hasil ini menunjukkan bahwa pemilihan threshold yang tepat berperan penting dalam memastikan sistem rekomendasi referensi jurnal untuk penentuan judul skripsi berbasis Cosine Similarity bekerja secara efektif dan akurat.

4. KESIMPULAN

Penelitian ini berhasil mengembangkan sistem rekomendasi referensi jurnal untuk penentuan judul skripsi dengan metode Cosine Similarity menggunakan data publikasi dari JATI ITN Malang. Sistem yang dibangun mampu memberikan rekomendasi judul skripsi yang relevan sesuai dengan kebutuhan pengguna. Evaluasi menunjukkan bahwa sistem memiliki tingkat Precision yang sangat baik dengan nilai 1.0 di seluruh percobaan, yang berarti seluruh rekomendasi dinilai tepat oleh model. Pengaruh threshold terbukti signifikan, di mana nilai yang lebih tinggi membuat rekomendasi semakin ketat namun menurunkan Recall dan Accuracy. Berdasarkan hasil pengujian, threshold optimal berada pada nilai 0.2 yang menghasilkan akurasi 88.89%, precision 100%, recall 83.33%, dan F1-score 90.91%.

Untuk meningkatkan performa sistem rekomendasi, pengembangan dapat dilakukan dengan mengombinasikan metode Cosine Similarity dengan teknik representasi teks lain seperti TF-IDF weighting,

Word2Vec, atau BERT embedding agar pemahaman konteks antar judul menjadi lebih baik. Penambahan variasi data dan masukan pengguna juga diperlukan agar cakupan rekomendasi semakin luas dan hasil yang diberikan lebih adaptif. Sistem dapat ditingkatkan dengan menyediakan visualisasi hasil evaluasi model sehingga pengguna dapat memahami performa sistem secara lebih informatif. Umpan balik pengguna juga penting untuk memperkuat validasi dan meningkatkan kualitas rekomendasi secara berkelanjutan.

REFERENSI

[1] R. Artikel and K. Kunci, "Jurnal Pendidikan Indonesia : SISTEM PENDETEKSI KEMIRIPAN JUDUL SKRIPSI BERBASIS WEB," vol. 5, no. 5, 2025, doi: 10.59818/jpi.v5i5.1964.

[2] L. Mayola, M. Hafizh, and D. M. Putra, "Perancangan Aplikasi *Similarity* Deteksi Kemiripan Judul Disertasi Berbasis Web," *J. Teknol. Dan Sist. Inf. Bisnis*, vol. 6, no. 2, pp. 452–257, 2024, doi: 10.47233/jteksis.v6i2.1164.

[3] R. K. Ari Shandy and G. N. Pramudyo, "Literasi Media Sosial Instagram Mahasiswa Ilmu Perpustakaan Universitas Diponegoro dalam Menentukan Informasi yang dipercaya," *Anuva J. Kaji. Budaya, Perpustakaan, dan Inf.*, vol. 7, no. 3, pp. 529–546, 2023, doi: 10.14710/anuva.7.3.529-546.

[4] D. Fisabilillah, K. Auliasari, and Y. A. Pranoto, "Sistem Rekomendasi Konversi Program Msib Dengan Mata Kuliah Prodi Informatika Itn Malang Menggunakan *Cosine similarity*," *SKANIKA Sist. Komput. dan Tek. Inform.*, vol. 8, no. 1, pp. 83–94, 2025, doi: 10.36080/skanika.v8i1.3325.

[5] N. Vendyansyah and Y. A. Pranoto, "Perancangan dan pembuatan aplikasi untuk mendeteksi kemiripan jawaban menggunakan cosine similarity," vol. 13, no. 1, pp. 23–28, 2021.

[6] Z. B. Toyibah and N. Aini, "Sistem Rekomendasi Skripsi Menggunakan Metode Content Based Filtering," *Infotech J. Technol. Inf.*, vol. 11, no. 1, pp. 89–98, 2025, doi: 10.37365/jti.v11i1.377.

[7] A. P. Muriaa, H. Sujainia, and H. S. Pratiwia, "Sistem Rekomendasi Artikel sebagai Acuan Studi Literatur Article Recommendation System as a Reference for Literature Studies Using the N-Gram Method," vol. 01, no. 1, pp. 69–84, 2023, doi: 10.26418/juristi.v1i1.61170.

[8] D. B. Elnursa, V. Nofriana, A. Syamsuri, and L. Cahyani, "Sistem Rekomendasi Pemilihan Program MSIB Bagi Mahasiswa Pendidikan Informatika," *J. Software, Hardw. Inf. Technol.*, vol. 3, no. 2, pp. 34–45, 2023, doi: 10.24252/shift.v3i2.92.

[9] A. A. Huda, R. Fajarudin, and A. Hadinegoro, "Sistem Rekomendasi Content-based Filtering Menggunakan TF-IDF Vector *Similarity* Untuk Rekomendasi Artikel Berita," *Build. Informatics, Technol. Sci.*, vol. 4, no. 3, pp. 1679–1686, 2022, doi: 10.47065/bits.v4i3.2511.

[10] Y. Januzaj and A. Luma, "*Cosine similarity* – A Computing Approach to Match *Similarity* Between Higher Education Programs and Job Market Demands Based on Maximum Number of Common Words," *Int. J. Emerg. Technol. Learn.*, vol. 17, no. 12, pp. 258–268, 2022, doi: 10.3991/ijet.v17i12.30375.

[11] T. P. Rinjeni, A. Indriawan, and N. A. Rakhmawati, "Matching Scientific Article Titles using *Cosine similarity* and Jaccard *Similarity* Algorithm," *Procedia Comput. Sci.*, vol. 234, no. 2023, pp. 553–560, 2024, doi: 10.1016/j.procs.2024.03.039.

[12] L. M. Purnomo and M. Ayub, "Analisis Data Hasil Web Scraping untuk Menentukan Kualitas Jurnal Ilmiah," *J. Strateg.*, vol. 3, no. 1, pp. 122–132, 2021.

[13] A. Apriani, H. Zakiyudin, and K. Marzuki, "Penerapan Algoritma *Cosine similarity* dan Pembobotan TF-IDF System Penerimaan Mahasiswa Baru pada Kampus Swasta," *J. Bumigora Inf. Technol.*, vol. 3, no. 1, pp. 19–27, 2021, doi: 10.30812/bite.v3i1.1110.

[14] M. Z. Mubarok, S. Rizqi, R. M. Rizal, A. B. N. Syalim, and N. Iksan, "Klasifikasi Sentimen Ulasan Aplikasi Claude Menggunakan Naive Bayes dan Support Vector Machine," *JUSIFOR J. Sist. Inf. dan Inform.*, vol. 4, no. 1, pp. 110–119, 2025, doi: 10.70609/jusifor.v4i1.7054.

[15] I. A. Fahrezi, Rudiman, and Nauval Azmi Verdikha, "Analisis Sentimen Twitter Atas Isu Hak Angket Menggunakan Pembobotan TF-IDF dan Algoritma SVM," *Sci-tech J.*, vol. 3, no. 2, pp. 179–192, 2024, doi: 10.56709/stj.v3i2.526.