Jurnal Informatika dan Rekayasa Perangkat Lunak ISSN 2656-2855
Vol. 7, No. 2, September 2025, Hal. 268-275 e-ISSN 2685-5518

Pemilihan Dosen Pembimbing Berdasarkan Judul Skripsi Mahasiswa menggunakan Metode *Cosine Similarity*

Ratri Andinisari[1*], **Fathur Riski**[2], **Karina Auliasari**[3]
[1] Teknik Sipil S1,Fakultas Teknik Sipil dan Perencanaan, Institut Teknologi Nasional Malang
[2,3] Teknik Informatika S1,Fakultas Teknologi Industri, Institut Teknologi Nasional Malang
[*]Email: aratri@lecturer.itn.ac.id, karina.auliasari@lecturer.itnac.id, riskifaturryu@gmail.com

Abstrak

Proses pemilihan dosen pembimbing yang selaras antara topik skripsi dan bidang keahlian dosen akan sangat membantu dalam proses penyusunan skripsi mahasiswa, namun pada kenyataannya, proses penentuan dosen pembimbing sering kali tidak didasarkan pada bidang keahlian dosen. Proses penentuan dosen pembimbing sering kali dilakukan secara manual dan kurang efisien, yang dapat menyebabkan ketidaksesuaian antara bidang keahlian dosen dengan topik skripsi mahasiswa. Dari permasalahan tersebut, pada penelitian ini dikembangkan pemodelan rekomendasi dalam pemilihan dosen pembimbing skripsi menggunakan metode Cosine Similarity, dengan menggunakan data judul skripsi dan daftar bidang keahlian dosen. Proses yang dimulai dengan melakukan prapemrosesan data teks pada judul skipsi, kemudian menghitung bobot kata menggunakan TF-IDF dan menghitung tingkat kemiripan untuk akhirnya mencocokkan antara judul skripsi mahasiswa dan bidang keashlian dosen yang dihitung menggunakan algoritma Cosine Similarity. Dari hasil pengujian proses mencocokkan topik penelitian dari judul skripsi mahasiswa dengan keahlian dan penelitian dosen dihasilkan secara akurat ini ditunjukkan oleh nilai precision sebesar 81% dan recall 90%, hasil ini juga menunjukkan bahwa tingkat relevansi rekomendasinya sangat tinggi. Dari hasil F1-Score yang diperoleh sebesar 85% menunjukkan bahwa pemodelan memiliki keseimbangan yang baik antara nilai presisi dan sensitivitas (recall).

***Kata kunci**: Pemodelan, Judul Skripsi, Dosen Pembimbing, Cosine Similarity, Keahlian*

Abstract

The process of selecting a thesis advisor that aligns the thesis topic with the advisor's area of expertise will greatly assist in the thesis preparation process for students. However, in reality, the process of determining the thesis advisor is often not based on the advisor's area of expertise. The process of determining the thesis advisor is often done manually and inefficiently, which can lead to a mismatch between the advisor's area of expertise and the student's thesis topic. From this problem, this research developed a recommendation model for selecting thesis supervisors using the Cosine Similarity method, utilizing thesis title data and the list of lecturers' areas of expertise. The process begins with preprocessing the text data of thesis titles, then calculating word weights using TF-IDF and measuring similarity to finally match the students' thesis titles with the lecturers' areas of expertise, which are calculated using the Cosine Similarity algorithm. From the results of testing the process of matching research topics from student thesis titles with the expertise and research of lecturers, it was produced accurately, as indicated by a precision value of 81% and a recall of 90%. This result also shows that the relevance level of the recommendations is very high. The F1 score obtained, which is 85%, indicates that the modeling has a good balance between precision and sensitivity (recall).

***Keyword :** Modeling, Thesis Title, Supervisor, Cosine Similarity, Expertise*

PENDAHULUAN

Pemilihan dosen pembimbing merupakan tahapan yang kritis dan merupakan langkah penting yang bisa menentukan lancar atau tidaknya skripsi mahasiswa. mahasiswa dapat memilih dosen pembimbing dengan cara pendekatan secara personal secara nmanual. karena proses manual seringkali membuat mahasiswa kesulitan menemukan dosen yang benar-benar sesuai dengan topik penelitian mahasiswa dan kerap menyebabkan pembimbing kurang optimal bahkan dapat memperpanjang masa penyeleaian skripsi mahasiswa tersebut karena ketidaktepatan judul skripsi mahasiswa tersebut dengan bidang keahlian dosen (Nasrullah, 2024) (Nashrullah, 2024). Sehingga, dapat berdampak pada mahasiswa dan juga dapat membebani dosen karena dengan topik yang diluar konteks kompetensi dosen, oleh karena itu diperlukan untuk solusi sistematis sistem rekomendasi yang akurat untuk membantu proses tersebut (Salam and Putraga Albahri, 2022).

Beberapa studi sebelumnya ada beberapa penelitian yang mencoba mengatasi dan mengusulkan sistem rekomendasi pembimbing dengan menggunakan metode pendekatan ada yang menggunakan *Multi-Class Support Vector Machine* dan Weighted product yang dimana digunakan untuk melakukan pencocokan latar belakang proposal skripsi dengan keahlian dosen. akan tetapi ada juga yang menggunakan metode *Collaborative Filtering* berbasis *Adjusted Cosine Similarity* yang diterapkan dalam sistem rekomendasi yang lain untuk mengukur kesamaan preferensi pengguna (Fisabilillah, Auliasari and Pranoto, 2025)

Penelitian ini dimana mengusulkan sistem rekomendasi berbasis *Cosine Similarity* yang digunakan untuk memetakan dan mengukur kemiripan kata kunci dalam judul skripsi mahasiswa dan bidang keahlian dosen. Kemampuannya dalam menganalisis kemiripan dokumen teks secara efisien, meskipun dengan jumlah data yang terbatas, menjadikan metode ini sebagai pilihan yang sesuai. Selain itu, implementasi metode ini cukup sederhana karena didukung oleh berbagai pustaka *Python* seperti *NumPy* dan *Scikit-Learn* yang memfasilitasi proses komputasi dan pemodelan data secara efisien (Dharmawan *et al.*, 2023). Sistem ini akan memproses beberapa tahapan dimulai dari tahapan *prepocessing text*, vektorisasi (misalnya menggunakan *TF-IDF*) dan juga perhitungan *Cosine Similarity* untuk menghasilkan daftar dosen dengan peringkat dan relevansi dosen pembimbing terbaik yang berdasarkan dengan topik skripsi mahasiswa (Wahyuni and Abdullah, 2025)

Diharapkan dengan sistem ini dapat memberikan rekomendasi dosen pembimbing yang lebih objektif dan akurat, serta mengurangi ketergantungan pada proses pemilihan dosen pembimbing secara manual dan dapat mempercepat bimbungan skripsi mahasiswa. Selain itu, hasil penelitian ini juga diharakan bisa menjadi dasar pengembangan sistem akademik berbasis teks dalam membantu mahasiswa untuk memilih dosen pembimbing skripsi yang sesuai dengan topik mahasiswa.

TINJAUAN PUSTAKA

2.1 Konsep dasar Sistem Rekomendasi

Sistem rekomendasi ini adalah sesuah sistem yang diguanakan untuk mempermudah dalam merekomendasikan dosen pembimbing untuk skripsi mahasiswa. Pendekatan ini dinilai sesuai untuk mencocokkan referensi judul skripsi mahasiswa dengan topik atau bidang penelitian dosen pembimbing (Wahyuni and Abdullah, 2025). *Cosine Similarity* dipilih sebagai metrik utama karena mampu mengubah representasi teks menjadi bentuk numerik, sehingga memungkinkan pengukuran tingkat kemiripan antar dokumen. Nilai kemiripan yang diperoleh dari perhitungan tersebut dapat dimanfaatkan untuk menentukan dosen yang paling relevan, berdasarkan tingkat kesesuaian yang diurutkan secara sistematis (Nasrullah, 2024)

Penelitian yang dilakukan (Herlinda et al. 2024) melakukan penelitian untuk mendeteksi kesamaan judul tugas akhir menggunakan TF-IDF dan *Cosine Similarity*. Prosesnya melibatkan *preprocessing* seperti *case folding*, *tokenizing*, *stopword removal*, dan *stemming*. Hasil menunjukkan *accuracy* sebesar 89,7 %, *precision* 72,4 %, serta *recall* 94,6 %—menunjukkan bahwa metode ini cukup andal dalam mengidentifikasi kemiripan antar dokumen judul akademik (Nasrullah, 2024).

Perbedaan penelitian ini dengan penelitian-penelitian tersebut adalah: (1) penggunaan dataset yang lebih luas dan terbaru (termasuk repositori publikasi dosen), (2) penggabungan fitur teks (judul, topik, abstrak) dengan profil publikasi dan bidang penelitian dosen, (3) evaluasi model menggunakan metrik lebih lengkap (*precision, recall, f-score*, dan mungkin evaluasi pengguna/koordinator skripsi), serta (4) implementasi sistem yang siap digunakan oleh mahasiswa/prodi dengan antarmuka serta kemampuan pembaruan data otomatis.

2.2 Metode Cosine Similarity

Metode *Cosine Similarity* adalah suatu teknik yang sering digunakan untuk menganalisis data dengan bentuk dokumen teks yang dimana digunakan untuk mengukur kemiripan teks yang membandingkan vektor untuk representasi dokumen teks untuk menentukan tingkat kesesuaiannya. Dalam konteks penelitian ini, sistem dikembangkan untuk memberikan rekomendasi dosen pembimbing yang relevan, dengan mempertimbangkan kesesuaian antara topik skripsi mahasiswa dan bidang keahlian dosen (Wahyuni and Abdullah, 2025).

Pemodelan ini dirancang untuk menyesuaikan topik skripsi mahasiswa dengan latar belakang keilmuan dosen, berdasarkan informasi dari hasil penelitian maupun publikasi yang relevan. Dimana, untuk menghitung nilai kemiripan konsinus tersebut dimana nilai tertinggi menunjukkan rank atau kecocokan terbaik antara judul skripsi mahasiswa dan dosen pembimbing. Selain itu, pendekatan ini mampu meningkatkan efisiensi dalam proses penentuan dosen pembimbing yang sebelumnya dilakukan secara manual, serta menghasilkan tingkat akurasi yang lebih optimal.

METODE

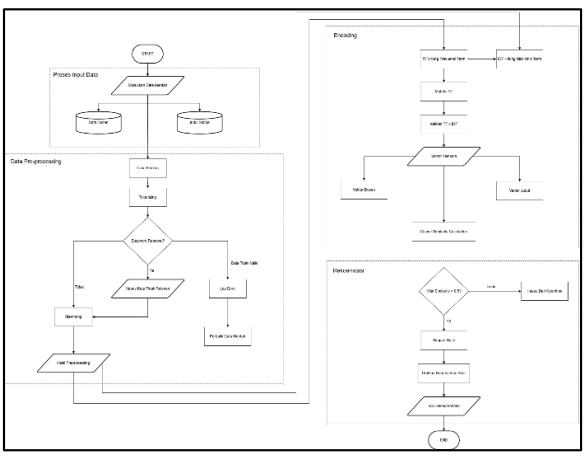
Metode penelitian ini bertujuan untuk membangun sebuah sistem rekomendasi dosen pembimbing yang dapat membantu mahasiswa dalam menentukan dosen pembimbing untuk skripsi mereka secara cepat, yang berdasarkan kesesuaian antara judul skripsi Melalui sistem ini, proses pencocokan antara topik skripsi mahasiswa dan pengalaman riset dosen, baik yang sedang berlangsung maupun yang telah dilakukan sebelumnya, dapat dilakukan secara otomatis. Implementasi sistem ini juga berpotensi mengurangi beban administratif dan waktu yang dibutuhkan oleh pihak kampus dalam menentukan dosen pembimbing secara manual. Dimana, pendekatan pada metode ini menggunakan metode *Cosine Similariry*, yaitu salah satu metode yang sering dan umum digunakan untuk mengukur kemiripan dan mencocokkan kemiripan antar dokumen yang berbasis teks, yang dimana sistem rekomendasi ini menggunakan pendekatan NLP dan algoritma *Cosine Similarity.*

3.1 Pengumpulan data

Dalam penelitian ini sumber utama data dalam penelitian terdiri dari dua sumber utama. Pertama, data judul skripsi mahasiswa dan data keahlian atau minat dosen. Informasi mengenai judul skripsi mahasiswa diperoleh dari data historis yang berisi daftar topik yang telah diajukan oleh mahasiswa pada tahun-tahun sebelumnya. Data ini digunakan sebagai representasi preferensi topik yang sering muncul dalam proses penyusunan tugas akhir yang disimpan dalam format excel dengan nama (data_judul_skripsi.xlsx). Data yang kedua yaitu data topik atau penelitian dosen, yang dimana data tersebut memberi gambaran atau sudut pandang mahasiswa terhadap dosen yang dimana menggambarkan bidang keilmuan yang ditekuni oleh masing-masing dosen. Data ini dikumpulkan melalui portofolio dosen, riwayat publikasi dan hasil rekap minat topik pembimbingan skripsi yang terdahulu atau yang pernah dilakukan oleh dosen tersebut. Dimana, data tersebut disimpan dalam format excel dengan nama (data_dosen_penelitian.xlsx). Seluruh data yang menjadi dasar dalam penelitian ini telah didokumentasikan dan dapat diakses secara publik melalui repositori GitHub berikut: https://github.com/FathurRiski08/Jurnal-cosine-similarity

3.2 Pengolahan Data

Sebelum masuk ke tahap perhitungan untuk kemiripan perhitungan *Cosine Similarity*, kita perlu melakukan pra-pemrosesan teks terlebih dahulu. tahap pra-pemrosesan adalah tahapan yang sangat penting dalam sistem rekomendasi berbasis teks, karena kualitas untuk representasi data sangat bergantung untuk kebersihan teks pada proses *input* teks. Dimana, tujuan dari pra-pemrosesan teks tersebut adalah untuk mengurangi *noise* pada teks dan menstandarkan kata tersebut dalam bentuk numerik agar bisa diproses secara numerik untuk menggunakan metode representasi vektor seperti *TF-IDF*. Adapun *Flowchart* dalam pengumpulan data tersebut yang tertera pada gambar 1.



Gambar 1. Alur Metode Penelitian.

Adapun tahapan untuk pra-pemrosesan datanya yaitu sebagai berikut :

a. *Input* Data.

Tahapan pertama yang dilakukan adalah yaitu menginputkan data yang diperlukan terlebih dahulu. Yaitu data judul skripsi dan data dosen penelitian, untuk dibandingkan dari kedua data tersebut. Kemudian setelah menginputkan data tersebut bisa memasuku pra-pemrosesan berikutnya (Fisabilillah, Auliasari and Pranoto, 2025).

b. *Case Floding.*

Pada tahapan ini yaitu pada proses *Case Floding,* yag dimana semua huruf yang berada didalam data teks yang kemudian dikonversi menjadi huruf kecil*(lowercase)* (Wahyuni and Abdullah, 2025). Dimana, hal ini digunakan untuk menghindari perbedaan misalnya, "Sistem Rekomendasi" menjadi "sistem rekomendasi".

c. *Tokenisasi*.

Tokenisasi adalah suatu proses dimana untuk memecah teks *input* menjadi bagian-bagian yang lebih kecil seperti katau atau frasa. Dimana, langkah ini digunakan untuk memisahkan kata

dalam kalimat disebuah dokumen yang digunakan sebagai *input*(Nasrullah, 2024). Langkah ini bertujuan untuk mengubah kata atau frasa menjadi bentuk vektor numerik, guna memudahkan sistem dalam melakukan perhitungan kemiripan antar teks.

d. *Stopword Removal*.

Stopword removal adalah proses dimana untuk menghilangkan kata-kata umum yang tidak memiliki makna secara signifikan atau tidak informatif, seperti "dan", "di" atau "yang"(Nasrullah, 2024). Dimana,langkah ini sangat penting karena bisa membantu mengurangi *noise* pada data dokumen.

e. *Stemming*.

Stemming adalah proses dimana sebuah kata tersebut diubah, yang sebelumnya masih mengandung kata imbuhan menjadi bentuk dasarnya. Misalnya "berjalan" dan "dialankan"menjadi "jalan"(Rinjeni, Indriawan and Rakhmawati, 2024).

f. *Normalisasi*

Normalisasi adalah tahapan proses yang digunakan untuk membakukan kata-kata tidak baku atau slag menjadi bentuk kata yang formal(Januzaj and Luma, 2022).

g. *Representasi* vektor TF-IDF.

Pada tahapan representasi vektor TF-IDF ini adalah suatu proses untuk mengubah teks menjadi bentuk numerik atau angka. Sehingga, agar bisa diproses oleh algoritma *machine learning.* Berikut adalah TF-IDF untuk persamaan menghitung bobot kata dengan menggunakan beberapa persamaan berikut .

1) *Term Frequency* (TF)

Digunakan untuk menghitung frekuensi kemunculan suatu kata*(term)* yang berada pada ssebuah dokumen. Dengan menggunkan persamaan (1).

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

2) *Inverse Document Frequency* (IDF)

Digunakan untuk mengukur seberapa jarang kata*(term)* tersebut muncul di seluruh koreksi dokumen. Dengan menggunkan persamaan (2).

$$\mathrm{IDF}(t,d) = \log\left(\frac{Total\ jumlah\ dokumen\ N}{Jumlah\ dokumen\ yang\ mengandung\ term\ t}\right) \quad (2)$$

Catatan, N adalah total dokumen yang berada didalam *corpus*, pembilangnya biasanya +1. Dengan menggunkan persamaan (3)

$$\mathrm{IDF}(t,d) = \log\left(\frac{N}{1+dokumen\ dengan\ term\ t}\right) + 1 \quad (3)$$

3) TF-IDF

Dimana TF-IDF adalah bobot akhir dari perkalian TF dan IDF. Dengan menggunkan persamaan (4).

$$\mathrm{TF\text{-}IDF}(t,d,D) = \mathrm{TF}(t,d) \times \mathrm{IDF}(t,D) \quad (4)$$

h. *Cosine Similarity*

Hasil dari perkalian TF-IDF dari data judul skripsi dan data dosen penelitian dibandingkan tingkat kesamaannya dengan menggunakan *Cosine Similarity. Cosine Similarity* mengukur kesamaan antara kedua vektor berdasarkan pada sudutnya, semakin kecil sudut *consinus* (mendekati 1), maka semakin tinggi juga kemiripannya(Fisabilillah, Auliasari and Pranoto, 2025). *Cosine Similarity* dijabarkan dengan menggunakan persamaan (5) berikut :

$$Similarity = cos(\theta) = \frac{A.B}{||A||||B||} = \frac{\sum_{i=1}^{n} A_i\ .\ B_i}{\sqrt{\sum_{i=1}^{n}(A_i)2}\ \sqrt{\sum_{i=1}^{n}(B_i)2}} \quad (5)$$

i. Rekomendasi *Output*

Hasil skor dari perhitungan *Cosine Similarity*, digunakan untuk merekomendasikan judul skripsi mahasiswa dengan dosen pembimbing. Dimana, hasil dari perhitungan ini disusun secara berurutan dari skor atau rank tertinggi(Fisabilillah, Auliasari and Pranoto, 2025). Judul skripsi dan nama dosen yang paling tinggi tingkat akurasinya berada pata rank teratas sebagai rekomendasi utama.

HASIL DAN PEMBAHASAN

Pada bagian ini akan dijelaskan hasil dan pembahasan dari penelitian tersebut yang sesuai dengan pembahasan tentang sistem rekomendasi dosen pembimbing :

4.1 Praprocessing Data

Data yang digunakan dalam penelitian ini terdiri dari dua set data yang disimpan dalam format *Excel*,yaitu: (1) data_judul_skripsi_mahasiswa.xlsx, dan (2) data_dosen_penelitian.xlsx.

Data dikumpulkan dalam bentuk teks mentah dan disimpan sebagai *file spreadsheet* maupun di dalam *database*. Tahap pra-analisis diawali dengan penyesuaian isi data untuk memastikan kualitas dan keteraturan. Proses ini melibatkan penyaringan entri yang duplikat dan penyesuaian format agar seluruh data berada dalam kondisi yang siap dianalisis secara sistematis.

Pada tahapan praprosesing data ini dilakukan proses pembersihan atau *cleaning* data yaitu dengan melakukan pembersihan dataset dari kolom-kolom yang tidak terpakai, karakter-karakter khusus yang tidak digunakan, dan juga data-data yang Redundansi terjadi ketika sebuah kata atau frasa digunakan lebih dari sekali dalam kalimat tanpa memberikan makna baru, sehingga menyebabkan pemborosan dalam penulisan. Berikut ini adalah hasil dari proses *praprocessing*:

```
# Load Data Dosen
file_path_dosen = "/content/drive/MyDrive/PAPPER JURNAL SINTA/data_dosen_penelitian.xlsx"
df_dosen = pd.read_excel(file_path_dosen)

df_dosen["Judul Penelitian"] = df_dosen["Judul Penelitian"].apply(preprocess_text)

# Load Data Skripsi Mahasiswa
file_path_skripsi = "/content/drive/MyDrive/PAPPER JURNAL SINTA/data_judul_skripsi.xlsx"
df_skripsi = pd.read_excel(file_path_skripsi)
df_skripsi["Judul Skripsi"] = df_skripsi["Judul Skripsi"].apply(preprocess_text)
```

Gambar 2. *Input data*

Pada gambar 2 menunjukkan kode tersebut merepresentasikan tahapan awal pra-pemrosesan data dalam pembangunan sistem rekomendasi dosen pembimbing berdasarkan judul skripsi mahasiswa. Proses dimulai dengan membaca data topik penelitian dosen dari *file Excel* bernama data_dosen_penelitian.xlsx, lalu kolom "Judul Penelitian" diproses menggunakan fungsi *preprocess_text*. Fungsi ini bekerja dengan melakukan normalisasi, menghapus karakter yang tidak relevan, menyaring *stopword*, dan mengubah kata menjadi bentuk dasar menggunakan *stemming*. Tahapan serupa dilakukan pula terhadap data judul skripsi mahasiswa dari *file* data_judul_skripsi.xlsx, di mana kolom "Judul Skripsi" juga diproses agar memiliki struktur teks yang seragam. Penyamaan format teks ini penting untuk mendukung proses pembandingan antar dokumen secara optimal menggunakan metode *Cosine Similarity* di tahap berikutnya.

```
Data Asli:  Implementasi Machine Learning untuk Deteksi Penyakit Diabetes
Case Folding:  implementasi machine learning untuk deteksi penyakit diabetes
Tokenization:  ['implementasi', 'machine', 'learning', 'untuk', 'deteksi', 'penyakit', 'diabetes']
Filtering:  ['implementasi', 'machine', 'learning', 'deteksi', 'penyakit', 'diabetes']
Stemming:  ['implementasi', 'machin', 'learn', 'deteksi', 'penyakit', 'diabet']
```

Gambar 3. Hasil *preprocessing data*

Pada gambar 3 teks awal yang dianalisis merupakan kalimat "Implementasi Machine Learning untuk Deteksi Penyakit Diabetes." Langkah pertama dalam *preprocessing* adalah *case folding*, yaitu proses mengubah seluruh huruf menjadi format huruf kecil. Selanjutnya, dilakukan tokenisasi dengan memecah kalimat menjadi kumpulan kata-kata tunggal atau token, seperti 'implementasi', 'machine', dan 'learning'. Langkah berikutnya adalah *filtering*, yaitu menghilangkan kata-kata yang tidak memiliki nilai penting secara semantik atau dikenal sebagai *stopword*, seperti kata 'untuk'. Untuk tahapan terakhir yaitu ada *stemming* digunakan untuk mengembalikan ke bentuk dasarnya. Contoh hasil tahapan *preprocessing* yang ditunjukkan pada Gambar 4.

Tahapan	Hasil
Data awal	Implemntasi Machine Learning untuk Deteksi Penyakit Diabetes
Case Folding	implemntasi machine learning untuk deteksi penyakit diabetes
Tokenization	['implemtasi','machine','learning','untuk','deteksi','penyakit','diabetes']
Filtering	['implemtasi','machine','learning','deteksi','penyakit','diabetes']
Stemming	['implemtasi','machin','learn','deteksi','penyakit','diabet']

Gambar 4. Contoh hasil *text preprocessing*

Setelah melalui tahapan pembersihan dan pemrosesan teks, hasil data kemudian diekspor ke dalam *file* berformat *excel* dengan nama hasil_rekomendasi_dosen.xlsx untuk mempermudah dokumentasi dan analisis lanjutan, seperti sebagian hasil *output* yang ditunjukkan pada Gambar 5.

No	Nama Mahasiswa	NIM	Judul Skripsi	Dosen Pembimbing
1	Ahmad Rifaldi	TI2021001	analisi sentimen komentar twitter metod naiv bay	Dr. Karina Putri
2	Budi Santoso	TI2021002	penerapan algoritma segmentasi pelanggan	Dr. Farhan Maulana
3	Citra Dewi	TI2021003	deteksi penyakit diabet decis tree	Dr. Ahmad Syarif
4	Dian Lestari	TI2021004	prediksi harga saham long memori lstm	Prof. Mega Sari
5	Eko Prasetyo	TI2021005	klasifikasi gambar convolut neural network cnn	Dr. Ahmad Syarif
6	Fajar Hidayat	TI2021006	sistem rekomendasi film collabor filter	Dr. Karina Putri
7	Gina Maulida	TI2021007	analisi kinerja jaringan komput smk negeri 1	Prof. Budi Santoso
8	Hendra Saputra	TI2021008	pembangunan aplikasi berbasi web	Dr. Erlina Widya
9	Indah Permata	TI2021009	penerapan augment realiti media pembelajaran biolog	Dr. Farhan Maulana
10	Joko Susilo	TI2021010	pengembangan chatbot informasi akademik nlp	Dr. Karina Putri

Gambar 5. Sebagian hasil output rekomendasi

Sesuai Gambar 5 yang menyajikan informasi mengenai sepuluh mahasiswa yang sedang menjalani tahap penulisan skripsi. Informasi yang ditampilkan pada tabel tersebut meliputi nama mahasiswa, NIM, judul skripsi dan nama dosen

pembimbing masing-masing. Fokus penelitian mahasiswa dalam skripsi umumnya mengarah pada pengembangan sistem berbasis data dan kecerdasan mesin. Beberapa mahasiswa tertarik meneliti dalam ranah *Natural Language Processing* (NLP). Contohnya adalah Ahmad Rifaldi yang mengangkat judul skripsi mengenai analisis sentimen pada komentar Twitter menggunakan algoritma *Naive Bayes*. Selain itu, terdapat pula Joko Susilo yang mengembangkan penelitian dalam bidang serupa. Berikutnya juga ada yang tertarik dengan topik *Machine Learning* juga cukup dominan diantaranya ada citra dewi mengenai deteksi diabetes dengan mtode *Decision Tree*, dan tentunya ada sebagian yang mengambil tema sistem rekomendasi seperti fajar hidayat dengan pengembangan aplikasi berbasis web. Dari sisi dosen pembimbing , ada beberapa dosenpembimbing yang membimbing lebih dari satu mahasiswa, seperti Dr.Karina putri dan Dr. Ahmad Syarif yang menunjukkan bahwa pengalaman mereka dalam membimbing skripsi dibidang teknologi dan kecerdasan buatan.

4.2 Uji dan Traning Model

Tahap ini menggunakan algoritma *machine learning* yang digunakan menghitung TF-IDF dan *Cosine Similarity*, dimana pada proses ini didapatkan hasil dari kedua perhitungan tersebut untuk menentukan nama dosen pembimbing yang didapatkan dari data dosen penelitian, dan kemudian dicocokkan dengan data judul skripsi mahasiswa sehingga mendapatkan nama dosen pembimbing yang sesuai dengan judul skripsi mahasiswa tersebut.

TF-IDF Matrix:

	algoritma	analisis	berbasis	content	cuaca	\
Dr. Ahmad Syarif	0.000000	0.000000	0.000000	0.000000	0.372225	
Prof. Budi Santoso	0.000000	0.000000	0.447214	0.447214	0.000000	
Dr. Citra Lestari	0.000000	0.388614	0.000000	0.000000	0.000000	
Judul Skripsi Mahasiswa	0.493386	0.000000	0.000000	0.000000	0.388991	

	dalam	filtering	learning	machine	media	\
Dr. Ahmad Syarif	0.47212	0.000000	0.372225	0.372225	0.000000	
Prof. Budi Santoso	0.00000	0.447214	0.000000	0.000000	0.000000	
Dr. Citra Lestari	0.00000	0.000000	0.000000	0.000000	0.388614	
Judul Skripsi Mahasiswa	0.00000	0.000000	0.388991	0.388991	0.000000	

	menggunakan	nlp	pada	penerapan	prediksi	\
Dr. Ahmad Syarif	0.000000	0.000000	0.000000	0.47212	0.372225	
Prof. Budi Santoso	0.000000	0.000000	0.000000	0.00000	0.000000	
Dr. Citra Lestari	0.306388	0.388614	0.388614	0.00000	0.000000	
Judul Skripsi Mahasiswa	0.388991	0.000000	0.000000	0.00000	0.388991	

	rekomendasi	sentimen	sistem	sosial
Dr. Ahmad Syarif	0.000000	0.000000	0.000000	0.000000
Prof. Budi Santoso	0.447214	0.000000	0.447214	0.000000
Dr. Citra Lestari	0.000000	0.388614	0.000000	0.388614
Judul Skripsi Mahasiswa	0.000000	0.000000	0.000000	0.000000

Gambar 6. Hasil TI-IDF Matrix

Gambar 6 menunjukkan matriks TF-IDF *(Term Frequency-Inverse Document Frequency)*, yang digunakan untuk mengukur seberapa signifikan suatu kata dalam sebuah dokumen relatif terhadap seluruh kumpulan dokumen yang dianalisis. Dokumen yang dianalisis terdiri dari judul-judul skripsi mahasiswa yang dibimbing oleh 3 dosen, yaitu Dr. Ahmad Syarif, Prof.Budi Santoso dan Dr. Citra Lestari. Matriks terdiri dari baris *(rows)* yaitu nama dosen (Dr. Ahmad Syarif, Prof. Budi Santoso, Dr. Citra Lestari) Setiap baris pada tabel tersebut merepresentasikan dokumen individual, seperti kumpulan penelitian dosen (contohnya milik Citra Lestari) maupun judul skripsi mahasiswa., seperti "algoritma", "analisis", "berbasis", "content", "cuaca", "dalam", "filtering", "learning", "machine", "media", "menggunakan", dan "nlp". Semakin tinggi nilainya, semakin relevan kata tersebut untuk dokumen itu.Nilai TF-IDF yang ditunjukkan ambar 4 untuk nilai 0.000000 menunjukkan kata tersebut tidak muncul atau tidak signifikan dalam dokumen, nilai yang kurang dari 0 (> 0) menunjukkan kata tersebut memiliki bobot penting, misalnya untuk pada Dr. Ahmad Syarif kata "cuaca" memiliki nilai 0.372225, menunjukkan bahwa kata ini relevan pada dokumen Dr. Ahmad Syarif, pada Prof. Budi Santoso kata "analisis" dan "berbasis" masing-masing memiliki nilai 0.447214, menandakan keduanya sangat penting dalam dokumennya, untuk Dr. Citra Lestari kata "media" memiliki nilai 0.388614, menunjukkan keterkaitan dengan topik yang dibahas dan Judul Skripsi Mahasiswa kata "algoritma" (0.493386) dan "prediksi" (0.388991) memiliki bobot tinggi, mengindikasikan fokus skripsi pada topik tersebut.Jika dianalisis dari dokumen Dr. Ahmad Syarif dokumennya berkaitan dengan "cuaca", "filtering", "learning", "machine", dan "prediksi", sebagai contoh topiknya sistem prediksi cuaca berbasis machine learning. Untuk Prof. Budi Santoso dokumennya fokus pada "analisis berbasis" dan "rekomendasi sistem", contoh topiknya analisis rekomendasi sistem berbasis algoritma tertentu. Untuk Dr. Citra Lestari dokumennya terkait "media", "nlp", "sentiment", dan "sosial", contoh topiknya analisis sentiment di media sosial menggunakan NLP.

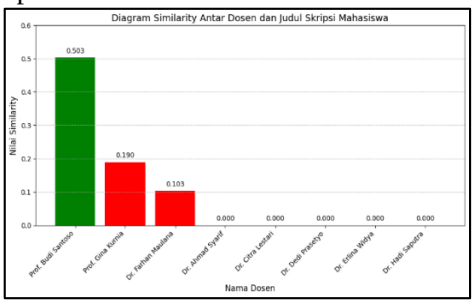
```
Rekomendasi Dosen Pembimbing berdasarkan judul skripsi:
Dr. Ahmad Syarif - Similarity: 0.5792
Dr. Citra Lestari - Similarity: 0.1192
Prof. Budi Santoso - Similarity: 0.0000
```

Gambar 7. Hasil *Cosine Similarity*

Berdasarkan hasil perhitungan *Cosine Similarity* pada Gambar 7. Dr. Ahmad Syarif merupakan dosen yang paling relevan untuk menjadi pembimbing dengan nilai kesesuaian sebesar 0,5792. Dr Citra Lesstari memiliki tingkat kesamaan 0,1192, Prof Budi Santoso memperoleh skor 0,0000, yang menunjukkan bahwa topik skripsi tidak sesuai dengan bidang keahlian yang dikuasainya. Oleh karena itu, rekomendasi dosen pembimbing utamanya adalah Dr. Ahmad Syarif.

4.3 Interpretasi Hasil

Pada tahap interpretasi hasil, dilakukan visualisasi dalam bentuk diagram blok untuk menampilkan hubungan antara nama dosen dan tingkat kesesuaian mereka terhadap judul skripsi mahasiswa.



Gambar 8. Hasil *Rekomendasi Dosen*

Grafik pada Gambar 8 tersebut menggambarkan derajat kesesuaian antar topik skripsi mahasiswa dan keahlian dosen masing-masing, berdasarkan hasil perhitungan tingkat kemiripan *(Similarity)* yang diperoleh. Dari hasil visualisasi prof dr budi santoso memiliki skor *similarity* tertinggi yaitu, 0,503 yang menandakan nahwa topik skripsi paling selaras dengan keahlian beliau. Diurutan selanjutnya ada prof gina kurnia memperoleh nilai 0.190 dan dr farhan maulana sebesar 0.103 yang masih menunjukkan tingkat relevansi meskipun lebih rendah. Sementara itu, beberapa deson seperti dr ahmad syarif dan dr citra lestari dan beberapa dosen yang lain memperoleh 0.0000, yang mengindikasikan tidak ada relevannya dengan keahlian mereka antara judul skripsi mahasiswa.

Gambar 9. Hasil *Word Cloud*

Wordcloud pada gambar 9 menunjukkan beberapa dafrat nama dosen yang paling sering direkomendasikan sebagai dosen pembimbing skripsi. Pada *Wordcloud* tersebut yang paling sering muncul adalah nama Budi Santoso dan tampil menonjol, yang menandakan frekuensi kemunculannya paling tinggi. Selain itu, nama-nama seperti gina,kurnia dan farhan juga cukup sering muncul. Begitu pula dengan sabaliknya nama-nama yang jarang muncul dalam frekuensi kemunculannya direkomendasi dosen pembimbing teksnya akan terlihat lebih kecil. Visualisai ini membantu mengidentifikasi doen yang paling relevan dengan topik skrpsi mahasiswa

SIMPULAN

Dari hasil pengujian proses mencocokkan topik penelitian dari judul skripsi mahasiswa dengan keahlian dan penelitian dosen dihasilkan secara akurat ini ditunjukkan oleh nilai *precision* sebesar 81% dan *recall* 90%, hasil ini juga menunjukkan bahwa tingkat relevansi rekomendasinya sangat tinggi. Dari hasil *F1-Score* yang diperoleh sebesar 85% menunjukkan bahwa pemodelan memiliki keseimbangan yang baik antara nilai presisi dan sensitivitas *(recall)*. Hasil ini mencerminkan bahwa pemodelan cukup efektif dalam memberikan rekomendasi dosen pembimbing yang relevan, serta dapat menjadi alat bantu yang efisien bagi pihak akademik dalam proses penugasan pembimbing efisien.

Sebagai penelitian lanjutan, disarankan untuk mengeksplorasi pendekatan berbasis pembelajaran mesin modern seperti BERT atau Sentence-BERT yang mampu menangkap makna semantik dari teks judul maupun abstrak skripsi. Selain itu, sistem rekomendasi juga dapat diperluas dengan mempertimbangkan faktor non-teknis, seperti beban bimbingan dan preferensi mahasiswa, sehingga hasil rekomendasi menjadi lebih akurat dan aplikatif. Implementasi dalam bentuk sistem berbasis web atau aplikasi mobile dengan evaluasi keterlibatan pengguna juga dapat dilakukan untuk mendukung penerapan di lingkungan universitas secara nyata.

DAFTAR PUSTAKA

Dharmawan, H. *et al.* (2023) *Hafizh Dharmawan, Sistem Rekomendasi Buku Dengan Metode K-Nearest Neighbor Pada Gramedia SISTEM REKOMENDASI BUKU DENGAN METODE K-NEAREST NEIGHBOR (K-NN) PADA GRAMEDIA*, *Sistem Informasi*.

Fisabilillah, D., Auliasari, K. and Pranoto, Y.A. (2025) 'Sistem Rekomendasi Konversi Program MSIB Dengan Mata Kuliah Prodi Informatika ITN Malang Menggunakan Cosine Similarity', *SKANIKA: Sistem Komputer dan Teknik Informatika*, 8(1), pp. 83–94.

Januzaj, Y. and Luma, A. (2022) 'Cosine Similarity – A Computing Approach to Match Similarity Between Higher Education Programs and Job Market Demands Based on Maximum Number of Common Words', *International Journal of Emerging Technologies in Learning*, 17(12), pp. 258–268. Available at: https://doi.org/10.3991/ijet.v17i12.30375.

Nasrullah, A.H. (2024) 'Integrasi Tf-Idf Dan Algoritma

Cosine Similarity Untuk Deteksi Tingkat Kemiripan Judul Penelitian (Studi Kasus Mahasiswa Fakultas Ilmu Komputer UNISAN Gorontalo)', *INTEC Journal: Information Technology Education Journal*, 3(3). Available at: https://scholar.google.com/,.

Rinjeni, T.P., Indriawan, A. and Rakhmawati, N.A. (2024) 'Matching Scientific Article Titles using Cosine Similarity and Jaccard Similarity Algorithm', in *Procedia Computer Science*. Elsevier B.V., pp. 553–560. Available at: https://doi.org/10.1016/j.procs.2024.03.039.

Salam, A. and Putraga Albahri, F. (2022) 'Sistem Rekomendasi Tugas Akhir Mahasiswa pada AMIK Indonesia untuk Mendukung Merdeka Belajar-Kampus Merdeka Menggunakan Metode Collaborative Filtering (CF)', *Jurnal Teknologi Informasi dan Komunikasi)*, 6(2), p. 2022. Available at: https://doi.org/10.35870/jti.

Wahyuni, S. and Abdullah, A. (2025) 'Penggunaan Information Retrieval untuk Mendeteksi Kesamaan Judul Skripsi dengan Modified Cosine Similarity', 8(2), pp. 117–126. Available at: https://doi.org/10.31764/justek.vXiY.ZZZ.

Dharmawan, H. *et al.* (2023) *Hafizh Dharmawan, Sistem Rekomendasi Buku Dengan Metode K-Nearest Neighbor Pada Gramedia SISTEM REKOMENDASI BUKU DENGAN METODE K-NEAREST NEIGHBOR (K-NN) PADA GRAMEDIA*, *Sistem Informasi*.

Fisabilillah, D., Auliasari, K. and Pranoto, Y.A. (2025) 'Sistem Rekomendasi Konversi Program MSIB Dengan Mata Kuliah Prodi Informatika ITN Malang Menggunakan Cosine Similarity', *SKANIKA: Sistem Komputer dan Teknik Informatika*, 8(1), pp. 83–94.

Januzaj, Y. and Luma, A. (2022) 'Cosine Similarity – A Computing Approach to Match Similarity Between Higher Education Programs and Job Market Demands Based on Maximum Number of Common Words', *International Journal of Emerging Technologies in Learning*, 17(12), pp. 258–268. Available at: https://doi.org/10.3991/ijet.v17i12.30375.

Nasrullah, A.H. (2024) 'Integrasi Tf-Idf Dan Algoritma Cosine Similarity Untuk Deteksi Tingkat Kemiripan Judul Penelitian (Studi Kasus Mahasiswa Fakultas Ilmu Komputer UNISAN Gorontalo)', *INTEC Journal: Information Technology Education Journal*, 3(3). Available at: https://scholar.google.com/,.

Rinjeni, T.P., Indriawan, A. and Rakhmawati, N.A. (2024) 'Matching Scientific Article Titles using Cosine Similarity and Jaccard Similarity Algorithm', in *Procedia Computer Science*. Elsevier B.V., pp. 553–560. Available at: https://doi.org/10.1016/j.procs.2024.03.039.

Salam, A. and Putraga Albahri, F. (2022) 'Sistem Rekomendasi Tugas Akhir Mahasiswa pada AMIK Indonesia untuk Mendukung Merdeka Belajar-Kampus Merdeka Menggunakan Metode Collaborative Filtering (CF)', *Jurnal Teknologi Informasi dan Komunikasi)*, 6(2), p. 2022. Available at: https://doi.org/10.35870/jti.

Wahyuni, S. and Abdullah, A. (2025) 'Penggunaan Information Retrieval untuk Mendeteksi Kesamaan Judul Skripsi dengan Modified Cosine Similarity', 8(2), pp. 117–126. Available at: https://doi.org/10.31764/justek.vXiY.ZZZ.