

## SISTEM PENGECEKAN FORMAT TUGAS AKHIR MENGGUNAKAN *RULE-BASED SYSTEM* DAN *SIMILARITY* JUDUL

Daililah Syahrina Aulia<sup>1\*</sup>, Sam Farisa Chaerul Haviana<sup>2</sup>

<sup>1,2</sup>Universitas Islam Sultan Agung, Indonesia

[32602200051@std.unissula.ac.id](mailto:32602200051@std.unissula.ac.id)<sup>1\*</sup>

[sam@unissula.ac.id](mailto:sam@unissula.ac.id)<sup>2</sup>

\*corresponding author

Received: 10-05- 2026

Revised: 21-05-2026

Approved: 28-05-2026

### ABSTRAK

Pemeriksaan format laporan Tugas Akhir secara manual membutuhkan waktu yang cukup lama dan berpotensi menimbulkan ketidakkonsistenan dalam proses evaluasi dokumen. Penelitian ini bertujuan untuk mengembangkan sistem pengecekan format laporan Tugas Akhir menggunakan metode Rule-Based System dan Similarity judul BAB berbasis TF-IDF serta Cosine Similarity. Sistem dikembangkan untuk membantu proses pemeriksaan kesesuaian format dan struktur penulisan laporan Tugas Akhir secara otomatis. Proses pengecekan meliputi struktur BAB, bagian awal dokumen, format penulisan, serta tingkat kemiripan struktur dokumen terhadap template standar. Sistem dibangun menggunakan bahasa pemrograman Python. Hasil evaluasi menunjukkan bahwa sistem mampu melakukan pengecekan format dokumen dengan baik menggunakan threshold 70 dan menghasilkan nilai accuracy sebesar 92,86%, precision sebesar 100%, dan recall sebesar 92,65%. Sistem juga dapat memberikan informasi kesalahan dan rekomendasi perbaikan dokumen secara otomatis sehingga membantu pengguna dalam memperbaiki laporan Tugas Akhir sesuai standar penulisan yang berlaku

**Kata kunci:** Cosine Similarity, Format Validation, Rule-Based System, TF-IDF

### PENDAHULUAN

Laporan Tugas Akhir merupakan salah satu syarat akademik yang wajib diselesaikan oleh setiap mahasiswa. Dalam proses penyusunannya, mahasiswa diwajibkan mengikuti pedoman penulisan yang telah ditetapkan agar laporan memiliki format yang seragam, sistematis, dan sesuai standar akademik (Patmawati, 2016) . Pedoman tersebut mencakup struktur penulisan, urutan BAB, format halaman, jenis huruf, ukuran huruf, spasi, serta kelengkapan bagian laporan seperti abstrak, daftar isi, daftar tabel, daftar gambar, dan daftar pustaka.

Namun dalam praktiknya, masih banyak mahasiswa yang mengalami kesalahan dalam penyusunan format laporan Tugas Akhir. Kesalahan tersebut meliputi ketidaksesuaian struktur BAB, penggunaan format penulisan yang tidak konsisten, kesalahan ejaan, margin, dan daftar pustaka. Penelitian yang dilakukan oleh Nofiandari menunjukkan bahwa masih ditemukan banyak kesalahan penulisan pada skripsi mahasiswa, baik dari segi struktur maupun penggunaan kaidah penulisan ilmiah. Selain itu, (Juli et al., 2024) juga menyatakan bahwa kesalahan penulisan skripsi mahasiswa sering terjadi pada format dokumen, struktur kalimat, tanda baca, dan daftar pustaka sehingga diperlukan proses evaluasi dokumen yang lebih sistematis dan konsisten. Oleh karena itu, diperlukan sistem otomatis yang mampu membantu proses pemeriksaan format laporan Tugas Akhir agar evaluasi dokumen dapat dilakukan secara lebih cepat, objektif, dan efisien.

Perkembangan teknologi informasi memungkinkan proses evaluasi dokumen dilakukan secara otomatis menggunakan sistem berbasis komputer. Salah satu pendekatan yang dapat digunakan adalah *Rule-Based System*, yaitu sistem yang

bekerja berdasarkan aturan tertentu yang telah ditentukan sebelumnya. Pendekatan *rule-based* dinilai sesuai untuk pengecekan format dokumen akademik karena aturan penulisan Tugas Akhir memiliki standar yang jelas dan terstruktur sehingga struktur penulisan lebih mudah dikenali dan dideteksi secara otomatis. (Juanda & Yadi, 2020).

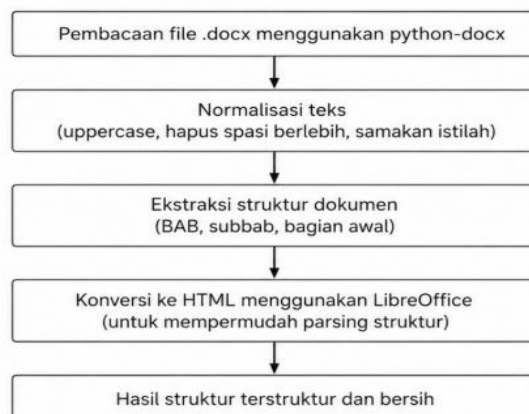
Selain pemeriksaan berbasis aturan, penelitian ini juga menerapkan metode *Similarity* menggunakan TF-IDF dan *Cosine Similarity* untuk mengukur tingkat kemiripan struktur dokumen dengan *template* laporan yang sesuai dengan buku panduan penulisan Tugas Akhir (Halim & Lasut, 2024). Metode TF-IDF digunakan untuk melakukan pembobotan kata, sedangkan *Cosine Similarity* digunakan untuk menghitung tingkat kemiripan antar dokumen berdasarkan struktur teks yang dimiliki (Asmaul Husnah Nasrullah, 2024). Penelitian sebelumnya menunjukkan bahwa metode TF-IDF dan *Cosine Similarity* efektif digunakan dalam pengukuran kemiripan dokumen teks dan sistem evaluasi dokumen akademik (Nilawati, Husaini and Salat, 2024).

Beberapa penelitian terdahulu telah membahas penerapan metode *Cosine Similarity* dalam proses deteksi kemiripan dokumen teks (Supiyanto & Sriyono, 2023). Metode tersebut dinilai efektif untuk mengukur tingkat kemiripan antar dokumen berdasarkan representasi teks yang dimiliki. Selain digunakan pada dokumen akademik, (Kurniadi et al., 2020) juga mengimplementasikan *Cosine Similarity* pada sistem arsip dokumen guna meningkatkan efektivitas pencarian berdasarkan kemiripan teks. Namun, penelitian-penelitian tersebut umumnya masih berfokus pada analisis *similarity* dokumen dan belum mengintegrasikan pemeriksaan format teknis dokumen dengan evaluasi struktur BAB secara otomatis. Oleh karena itu, penelitian ini mengembangkan sistem pengecekan format laporan Tugas Akhir dengan mengombinasikan *Rule-Based System* dan *Similarity* untuk menghasilkan proses evaluasi dokumen yang lebih menyeluruh, otomatis, dan konsisten.

## METODE PENELITIAN

### *Ekstraksi Dokumen*

Ekstraksi dokumen merupakan tahapan mempersiapkan data dokumen sebelum masuk ke proses evaluasi sistem. Tahap ini akan menuntut sistem untuk membaca dokumen, menormalisasi teks, ekstraksi struktur dokumen, serta konversi dokumen ke format HTML. Adapun gambar alur proses ekstraksi dokumen seperti pada gambar 1 di bawah ini.



**Gambar 1.** Alur ekstraksi dokumen

Pada gambar alur tersebut menunjukkan tahapan yang terjadi di ekstraksi data.

1. Pembacaan file .docx menggunakan *python-docx*, tahapan ini merupakan proses pembacaan dokumen menggunakan *library python-docx* untuk mengambil isi dan elemen dokumen.
2. Normalisasi teks, hasil pembacaan akan dinormalisasi dengan mengubah huruf menjadi *uppercase*, menghapus spasi berlebih, serta menyamakan beberapa istilah supaya format teks menjadi lebih konsisten.
3. Ekstraksi teks, sistem akan mengidentifikasi struktur penting pada dokumen seperti BAB, subbab, bagian awal laporan Proses tersebut dilakukan untuk mendukung evaluasi format dokumen dan perhitungan similarity judul BAB.
4. Konversi ke format HTML, dokumen dikonversi menjadi format HTML menggunakan *LibreOffice* untuk mempermudah proses parsing dan pembacaan struktur dokumen secara lebih terstruktur.

### Rule-Based System

*Rule-Based System* merupakan sistem berbasis aturan yang bekerja menggunakan logika if-then untuk menghasilkan keputusan berdasarkan aturan yang telah ditentukan sebelumnya. Sistem ini banyak digunakan dalam proses validasi dokumen karena mampu melakukan pemeriksaan secara konsisten dan objektif. *Rule-Based System* efektif digunakan pada sistem evaluasi dokumen karena mampu menerapkan aturan secara terstruktur dan mudah dikembangkan. Selain itu, *Rule-Based System* mampu melakukan proses ekstraksi dan pengelompokan informasi dokumen secara otomatis dengan tingkat akurasi yang tinggi (Ardiansyah et al., 2023).

**Tabel 1.** Komponen pemeriksaan laporan tugas akhir

Komponen Pemeriksaan	Aturan yang Digunakan
Struktur BAB dan subbab	Mengikuti sistematika laporan Tugas Akhir
Bagian awal dokumen	Cover, halaman judul, abstrak, daftar isi, daftar tabel, daftar gambar, dan lampiran
Margin dokumen	Atas 4 cm, kiri 4 cm, bawah 3 cm, kanan 3 cm
Jenis font	Times New Roman
Ukuran font	12 pt
Spasi penulisan	1,5
Format abstrak dan abstract	Sesuai pedoman penulisan
Format daftar pustaka	Menggunakan sistem Harvard
Sistematika BAB	BAB I sampai BAB V
Similarity struktur BAB	Dibandingkan dengan template standar
Format sitasi	Mengikuti pedoman penulisan ilmiah
Konsistensi struktur dokumen	Sesuai Buku Panduan Teknik Informatika UNISSULA

### TF-IDF

TF-IDF (Term Frequency – Inverse Document Frequency) merupakan

metode pembobotan kata yang digunakan untuk mengetahui tingkat kepentingan suatu kata dalam dokumen. Metode ini bekerja dengan menghitung frekuensi kemunculan kata pada dokumen dan membandingkannya dengan frekuensi kata pada seluruh dokumen. Kata yang sering muncul pada suatu dokumen namun jarang muncul pada dokumen lain akan memiliki bobot lebih tinggi. Menurut penelitian (Choirul Hana & Kurniawan, 2025) TF-IDF banyak digunakan dalam text mining dan information retrieval karena mampu menghasilkan representasi dokumen yang efektif.

$$TF - IDF(t, d) = TF(t, d) \times IDF(t) \quad (1)$$

Keterangan :

$TF-IDF(t, d)$  : bobot kata pada dokumen

$TF(t, d)$  : jumlah seluruh *term* pada dokumen

$IDF(t)$  : nilai inverse document *frequency*

$t$  : *term* atau kata

$d$  : dokumen

Berdasarkan rumus tersebut, nilai TF-IDF diperoleh dari hasil perkalian antara nilai term frequency (TF) dan inverse document frequency (IDF). Nilai TF digunakan untuk menghitung frekuensi kemunculan suatu kata pada dokumen, sedangkan nilai IDF digunakan untuk mengetahui tingkat pentingnya kata berdasarkan kemunculannya pada seluruh dokumen. Semakin tinggi nilai TF-IDF suatu kata, maka kata tersebut dianggap semakin penting dalam merepresentasikan isi dokumen. Penelitian yang dilakukan oleh (Riyani et al., 2019) menyatakan bahwa metode TF-IDF mampu menghasilkan pembobotan kata yang efektif dalam proses analisis kemiripan dokumen dan dapat digunakan bersama metode *Cosine Similarity* untuk menghitung tingkat kemiripan teks dokumen.

### ***Cosine Similarity***

*Cosine Similarity* merupakan metode yang digunakan untuk menghitung tingkat kemiripan antara dua dokumen berdasarkan representasi vektor. Metode ini mampu memberikan hasil pengukuran kemiripan dokumen yang cukup baik dan efektif digunakan dalam analisis dokumen teks (Pratama et al., 2019) . Perhitungannya dilakukan dengan mencari nilai cosinus sudut antara dua vektor dokumen. Nilai *Similarity* berada pada rentang 0 hingga 1, di mana nilai yang semakin mendekati 1 menunjukkan tingkat kemiripan yang semakin tinggi.

$$Cosine(A, B) = \frac{\{A \cdot B\}}{\{|A| \times |B|\}} \quad (2)$$

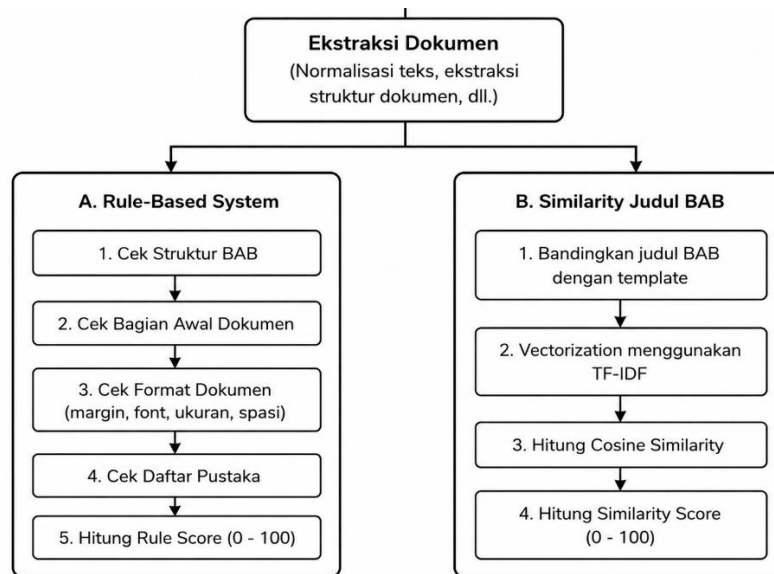
Keterangan :

$(A, B)$  : jumlah kemunculan term  $t$  pada dokumen  $d$

$|A|$  : panjang vektor A

$|B|$  : panjang vektor B

### ***Pemeriksaan Format Dokumen dengan Rule-Based System dan Similarity***



**Gambar 2.** Alur pemeriksaan format

Pemeriksaan format dokumen dilakukan untuk memeriksa kesesuaian format secara otomatis dengan menggunakan metode *Rule-Based System* dan *Similarity* yang merupakan gabungan dari *TF-IDF* dan *Cosine Similarity*. *Rule-Based System* digunakan untuk memeriksa elemen format dokumen berdasarkan buku panduan penulisan tugas akhir, sedangkan *Similarity* digunakan untuk menghitung tingkat kemiripan struktur terhadap template standar laporan tugas akhir.

#### Score Gabungan

Hasil proses pada Gambar 2 menghasilkan skor evaluasi dokumen dengan bobot sebesar 70% untuk *Rule-Based System* dan 30% untuk *Similarity*. Penggabungan nilai dilakukan untuk menghasilkan evaluasi yang seimbang antara kesesuaian format teknis dan kesesuaian struktur dokumen terhadap template standar penulisan tugas akhir. Pembobotan sebesar 70% untuk *Rule-Based System* dan 30% untuk *Similarity* digunakan karena pemeriksaan kesesuaian format teknis dokumen dianggap lebih penting dibandingkan tingkat kemiripan struktur judul BAB. *Rule-Based System* memiliki peran utama dalam memeriksa elemen wajib pada laporan seperti margin, jenis font, ukuran font, spasi, dan struktur dokumen sesuai pedoman penulisan Tugas Akhir. Sementara itu, metode *Similarity* digunakan sebagai pendukung untuk mengevaluasi kemiripan struktur BAB dan subbab terhadap *template* standar. Oleh karena itu, proporsi bobot yang lebih besar diberikan pada *Rule-Based System* agar hasil evaluasi lebih sesuai dengan kebutuhan validasi format dokumen akademik.

#### Evaluasi

$$Accuracy = \frac{\{TP + TN\}}{\{TP + TN + FP + FN\}} \quad (3)$$

$$Precision = \frac{TP}{TP + FP} \quad (4)$$

$$Recall = \frac{TP}{TP + FN} \quad (5)$$

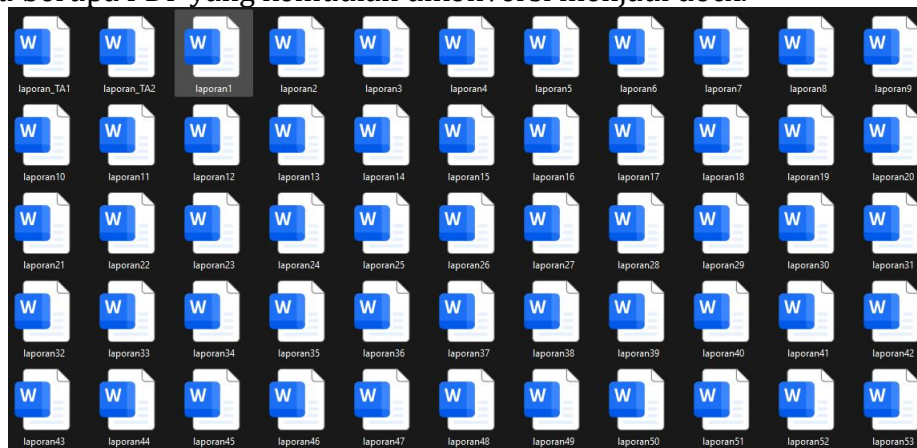
Evaluasi sistem dilakukan menggunakan confusion matrix untuk mengukur performa klasifikasi dokumen sesuai dan tidak sesuai terhadap template standar laporan Tugas Akhir. Confusion matrix terdiri dari *True Positive*(TP), *False Positive*(FP), *True negative* (TN), dan *False negative* (FN).

*True Positive*(TP) menunjukkan jumlah dokumen yang sesuai format dan berhasil diklasifikasikan sesuai oleh sistem. *False Positive*(FP) menunjukkan jumlah dokumen yang sebenarnya tidak sesuai namun diklasifikasikan sesuai oleh sistem. *True negative* (TN) menunjukkan jumlah dokumen tidak sesuai yang berhasil dikenali dengan benar oleh sistem. Sedangkan *False negative* (FN) menunjukkan jumlah dokumen yang sebenarnya sesuai namun diklasifikasikan tidak sesuai oleh sistem.

Berdasarkan nilai confusion matrix tersebut, dilakukan perhitungan accuracy, precision, dan recall untuk mengetahui tingkat performa sistem dalam melakukan evaluasi format laporan Tugas Akhir.

## HASIL DAN PEMBAHASAN

Dataset yang digunakan dalam penelitian ini diperoleh dari *repository Unissula* berupa PDF yang kemudian dikonversi menjadi docx.



Gambar 3. Kumpulan file laporan dengan format docx

## Pembacaan File

### BAB 1 PENDAHULUAN

#### 1.1 Latar Belakang

Bahasa Indonesia adalah salah satu bahasa dengan jumlah penutur terbanyak di dunia, digunakan oleh lebih dari 200 juta orang sebagai penutur asli dan total sekitar 270 juta orang. Namun perkembangan teknologi pemrosesan bahasa alami (*Natural Language Processing*) untuk Bahasa Indonesia masih jauh tertinggal dibandingkan bahasa sumber daya yang lebih banyak, seperti bahasa Inggris atau Mandarin. Salah satu hambatan utama dalam pengembangan teknologi ini adalah keterbatasan data terstruktur, terutama korpus paralel yang berisi pasangan kalimat benar-salah, yang sangat penting dalam tugas koreksi kesalahan tata bahasa otomatis (*Grammatical Error Correction*) (Musyafa dkk., 2022).

#### Gambar 4. Laporan sebelum diekstraksi

Pada gambar 4 menunjukkan tampilan dokumen asli sebelum dilakukan ekstraksi, sedangkan gambar 5 merupakan hasil pembacaan isi dokumen menggunakan *library python-docx*. Sistem berhasil mengekstraksi isi dokumen menjadi bentuk teks sehingga struktur dokumen seperti BAB, subbab, dan isi paragraf dapat diproses pada tahap ekstraksi dan evaluasi sistem.

BAB 1  
PENDAHULUAN

Latar Belakang

Bahasa Indonesia adalah salah satu bahasa dengan jumlah penutur terbanyak Di lingkungan perguruan tinggi, khususnya di program Studi Teknik Informa Untuk mengatasi masalah tersebut, diperlukan sistem pengoreksian otomatis Adaptation (LoRA), model ini dapat diadaptasi ke laporan praktikum secara Meskipun sudah ada beberapa penelitian GEC untuk bahasa Indonesia, hampir Penelitian ini bertujuan memberikan solusi praktis bagi mahasiswa dan dos

**Gambar 5.** Hasil pembacaan sistem

**Normalisasi Teks**

Pada tahap normalisasi, sistem melakukan pembersihan dan penyamaan format teks hasil pembacaan dokumen agar lebih mudah diproses pada tahap evaluasi sistem. Proses ini dilakukan karena hasil ekstraksi dokumen .docx sering memiliki perbedaan format penulisan, seperti penggunaan huruf kecil dan besar yang tidak konsisten serta adanya spasi berlebih pada teks.

SEBELUM NORMALISASI:	SETELAH NORMALISASI:
Bab 1	BAB 1
pendahuluan	PENDAHULUAN
Latar Belakang	LATAR BELAKANG

**Gambar 6.** Before after normalisasi

Tahap pertama yaitu mengubah seluruh teks menjadi huruf uppercase agar sistem dapat mengenali struktur dokumen dengan lebih konsisten. Sebagai contoh, teks “Bab 1” diubah menjadi “BAB 1” dan “pendahuluan” menjadi “PENDAHULUAN”. Tahap berikutnya adalah menghapus spasi berlebih pada teks. Proses ini dilakukan untuk merapikan struktur tulisan sehingga teks yang memiliki jarak tidak beraturan dapat menjadi lebih terstruktur. Sebagai contoh, teks “Latar Belakang” diubah menjadi “LATAR BELAKANG”.

**Ekstraksi Teks**

Heading 1 : KATA PENGANTAR  
Heading 1 : DAFTAR ISI  
Heading 1 : DAFTAR GAMBAR  
Heading 1 : DAFTAR TABEL  
Heading 1 : ABSTRAK  
Heading 1 :  
Heading 1 : BAB 1  
PENDAHULUAN  
Heading 2 : Latar Belakang  
Heading 2 : Perumusan Masalah  
Heading 2 : Pembatasan Masalah  
Heading 2 : Tujuan  
Heading 2 : Manfaat  
Heading 2 : Sistematika Penulisan  
Heading 1 :  
Heading 1 : BAB II  
TINJAUAN PUSTAKA DAN DASAR TEORI  
Heading 2 : Tinjauan Pustaka

**Gambar 7.** Hasil ekstraksi dokumen

Pada gambar 7 menunjukkan hasil ekstraksi struktur dokumen yang

dilakukan sistem menggunakan heading pada file .docx. Sistem berhasil mengenali bagian awal laporan, struktur BAB, dan subbab secara otomatis berdasarkan jenis heading yang terdapat pada dokumen. Hasil ekstraksi menunjukkan bahwa heading utama seperti BAB I dan BAB II terdeteksi sebagai Heading 1, sedangkan subbab seperti Latar Belakang dan Perumusan Masalah terdeteksi sebagai Heading 2. Struktur dokumen yang berhasil diekstraksi kemudian digunakan pada proses *Rule-Based System* dan *Similarity* judul BAB.

### **Hasil Rule Based Similarity**

Berdasarkan hasil pemeriksaan, sistem berhasil mendeteksi struktur dokumen, bagian awal laporan, isi BAB, serta format dokumen seperti margin, font, ukuran font, spasi, dan daftar Pustaka. Tanda centang menunjukkan bahwa komponen dokumen telah sesuai dengan aturan format yang ditentukan. Hal ini dapat dilihat pada gambar 8 di bawah ini.

```
=====
RULE CHECK : laporan_TA1.docx
=====
{'Struktur': 100.0, 'Bagian Awal': 100.0, 'Urutan BAB': 100, 'BAB I': 100.0,

Rule Score :
100.0
```

**Gambar 8.** Hasil pengecekan *Rule-Based System*

### **Hasil Similarity**

Sistem melakukan perhitungan tingkat kemiripan struktur BAB dan subbab dokumen terhadap template standar laporan Tugas Akhir. Proses *Similarity* dilakukan menggunakan metode TF-IDF dan *Cosine Similarity*.

```
=====
SIMILARITY : laporan50.docx
=====
INFO: convert C:\TA DAILILAH\tugas akhir\ta\dataset\laporan50.docx
Overwriting: C:\TA DAILILAH\tugas akhir\ta\temp_html\laporan50.htm

WARNING: Could not find platform independent libraries <prefix>

Similarity Score : 61.96
```

**Gambar 9.** Hasil pengecekan *Similarity*

Gambar 9 tersebut menunjukkan hasil proses *Similarity* judul BAB pada sistem evaluasi format laporan Tugas Akhir. Pada tahap ini, sistem melakukan perbandingan struktur BAB dan subbab dokumen terhadap template standar laporan Tugas Akhir menggunakan metode TF-IDF dan *Cosine Similarity*.

### **Score Gabungan**

Nilai akhir (final score) diperoleh dari kombinasi kedua metode dengan pembobotan 70% untuk rule score dan 30% untuk *Similarity* score. Hasil penggabungan score digunakan oleh sistem untuk menentukan status dokumen, apakah termasuk sesuai atau tidak sesuai dengan template standar yang digunakan.

```
laporan50.docx
Similarity : 61.96
Rule Score : 92.8
Final Score : 83.548
INFO: convert C:\TA DAILILAH\tugas akhir\ta\dataset\laporan51.dc
Overwriting: C:\TA DAILILAH\tugas akhir\ta\temp_html\laporan51.f

WARNING: Could not find platform independent libraries <prefix>
```

**Gambar 10.** Hasil penggabungan skor

### **Evaluasi**

Berdasarkan hasil evaluasi sistem, diperoleh nilai *True Positive*(TP) sebanyak 63 dokumen, *False Positive*(FP) sebanyak 0 dokumen, *True negative* (TN) sebanyak 2 dokumen, dan *False negative* (FN) sebanyak 5 dokumen. Nilai TP menunjukkan jumlah dokumen yang sesuai format dan berhasil diklasifikasikan dengan benar oleh sistem, sedangkan nilai TN menunjukkan jumlah dokumen tidak sesuai yang berhasil dikenali secara tepat. Nilai FP sebesar 0 menandakan bahwa tidak terdapat dokumen tidak sesuai yang salah diklasifikasikan sebagai dokumen sesuai. Sementara itu, nilai FN sebesar 5 menunjukkan masih terdapat beberapa dokumen sesuai yang belum berhasil dikenali secara tepat oleh sistem.

```
TP : 63
FP : 0
TN : 2
FN : 5
```

=== HASIL EVALUASI SISTEM ===

```
Accuracy : 92.86 %
Precision : 100.0 %
Recall : 92.65 %
```

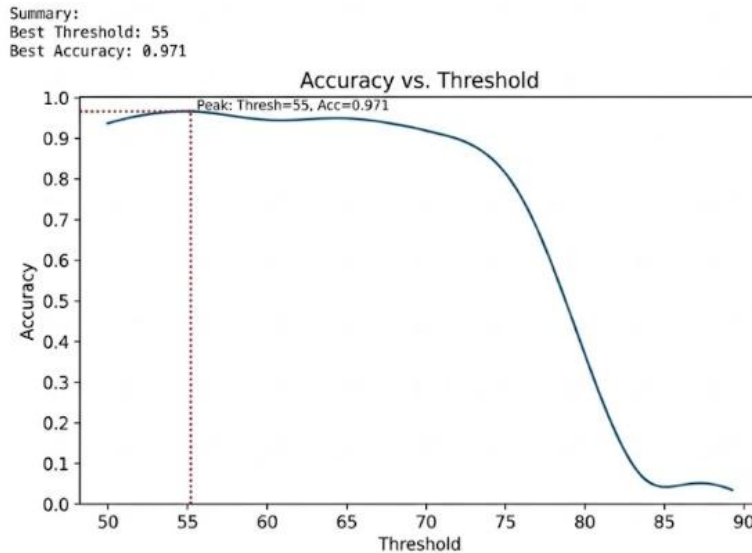
**Gambar 11.** Hasil evaluasi

Hasil pengujian menunjukkan bahwa sistem memperoleh nilai *accuracy* sebesar 92,86%, *precision* sebesar 100%, dan *recall* sebesar 92,65%. Nilai *accuracy* menunjukkan tingkat ketepatan sistem dalam melakukan klasifikasi dokumen terhadap seluruh data uji. Nilai *precision* sebesar 100% menandakan bahwa seluruh dokumen yang diprediksi sesuai oleh sistem benar-benar merupakan dokumen yang sesuai, sehingga tidak ditemukan kesalahan pada prediksi positif. Adapun nilai *recall* sebesar 92,65% menunjukkan bahwa sistem mampu mengenali sebagian besar dokumen sesuai dengan baik, meskipun masih terdapat beberapa dokumen yang belum terdeteksi secara sempurna.

Nilai *False Negative* (FN) sebesar 5 menunjukkan bahwa masih terdapat beberapa dokumen yang sebenarnya sesuai format namun diklasifikasikan sebagai tidak sesuai oleh sistem. Kondisi ini kemungkinan disebabkan oleh adanya perbedaan kecil pada struktur heading, penggunaan format penulisan yang tidak sepenuhnya konsisten, atau ketidaksesuaian penulisan subbab terhadap template standar yang digunakan sistem. Selain itu, proses ekstraksi dokumen docx juga dapat memengaruhi hasil evaluasi ketika *heading* atau format tertentu tidak terbaca secara sempurna oleh sistem. Hal tersebut menyebabkan beberapa dokumen sesuai memperoleh nilai *similarity* atau *rule score* di bawah *threshold*

yang telah ditentukan. Temuan ini menunjukkan bahwa pengembangan sistem selanjutnya perlu difokuskan pada peningkatan proses ekstraksi dokumen dan penyesuaian aturan evaluasi agar sistem lebih toleran terhadap variasi format penulisan yang masih dapat diterima.

### Hasil Threshold



**Gambar 12.** Nilai *threshold*

Berdasarkan hasil pengujian *threshold*, sistem menunjukkan performa *accuracy* yang berbeda pada setiap nilai *threshold* yang digunakan. *Threshold* 55 menghasilkan *accuracy* tertinggi sebesar 97,1%, sedangkan *threshold* 50, 60, dan 65 masing-masing menghasilkan *accuracy* sebesar 95,7%. Pada *threshold* 70, sistem memperoleh *accuracy* sebesar 92,9% yang masih menunjukkan performa sangat baik dalam proses klasifikasi dokumen.

Namun, ketika nilai *threshold* ditingkatkan menjadi 75, 80, 85, dan 90, nilai *accuracy* mengalami penurunan yang cukup signifikan. Hal ini menunjukkan bahwa semakin tinggi *threshold* yang digunakan, maka sistem menjadi lebih ketat dalam menentukan kategori dokumen sesuai, sehingga jumlah dokumen yang diklasifikasikan sebagai tidak sesuai menjadi lebih banyak.

Meskipun *threshold* 55 menghasilkan *accuracy* tertinggi, penelitian ini menetapkan *threshold* 70 sebagai *threshold* final. Pemilihan tersebut dilakukan karena *threshold* 70 dinilai lebih sesuai dengan standar evaluasi format laporan Tugas Akhir serta tetap mampu memberikan performa sistem yang baik dan konsisten.

## KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, sistem pengecekan kesesuaian format laporan Tugas Akhir berhasil diimplementasikan menggunakan kombinasi metode *Rule-Based System* dan *Similarity* judul BAB berbasis *TF-IDF* serta *Cosine Similarity*. Sistem mampu melakukan pembacaan dokumen .docx, normalisasi teks, ekstraksi struktur dokumen, hingga evaluasi kesesuaian format laporan secara otomatis.

*Rule-Based System* berhasil digunakan untuk memeriksa komponen format dokumen seperti struktur BAB, subbab, margin, jenis *font*, ukuran *font*, spasi, dan daftar pustaka berdasarkan aturan yang telah ditentukan. Selain itu, metode *TF-IDF* dan *Cosine Similarity* mampu menghitung tingkat kemiripan struktur BAB dan

subbab dokumen terhadap template standar laporan Tugas Akhir dengan baik.

Hasil evaluasi menunjukkan bahwa sistem memperoleh nilai *accuracy* sebesar 92,86%, *precision* sebesar 100%, dan *recall* sebesar 92,65%. Nilai tersebut menunjukkan bahwa sistem memiliki tingkat ketepatan yang tinggi dalam mengklasifikasikan dokumen sesuai maupun tidak sesuai terhadap template standar yang digunakan. Pengujian *threshold* juga menunjukkan bahwa *threshold* 70 memberikan performa yang stabil dan dinilai paling sesuai untuk proses evaluasi format laporan Tugas Akhir.

Implementasi sistem ini juga memberikan kontribusi praktis bagi institusi pendidikan dalam membantu proses pemeriksaan format laporan Tugas Akhir secara lebih cepat, konsisten, dan objektif dibandingkan pemeriksaan manual. Selain meningkatkan efisiensi waktu proses evaluasi dokumen, sistem ini juga dapat membantu mengurangi beban kerja penelaah atau evaluator format laporan Tugas Akhir.

Namun, sistem yang dikembangkan masih memiliki keterbatasan pada jenis dokumen input yang hanya mendukung format .docx serta kemampuan pembacaan struktur dokumen yang masih bergantung pada penggunaan heading dan format penulisan yang konsisten. Oleh karena itu, penelitian selanjutnya dapat difokuskan pada pengembangan sistem agar mampu mendukung lebih banyak format dokumen serta meningkatkan kemampuan ekstraksi struktur dokumen yang lebih fleksibel terhadap variasi format penulisan laporan Tugas Akhir.

#### **DAFTAR PUSTAKA**

- Ardiansyah, M., Hantoro, K., & Rasim. (2023). IMPLEMENTASI RULE BASED SYSTEM UNTUK SISTEM JADWAL PAKAN IKAN KOMET OTOMATIS BERBASIS ANDROID. *Jurnal Informatika Dan Teknik Elektro Terapan*, 11(3). <https://doi.org/10.23960/jitet.v11i3.3243>
- Asmaul Husnah Nasrullah. (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), 113–118.
- Choirul Hana, R., & Kurniawan, D. (2025). Implementasi TF-IDF N-Gram dan Algoritma Nearest Centroid untuk Klasifikasi Topik Tugas Akhir. *Technology and Science (BITS)*, 7(3). <https://doi.org/10.47065/bits.v7i3.8859>
- Halim, J., & Lasut, D. (2024). Document Plagiarism Detection Application Using Web-Based TF-IDF and Cosine Similarity Methods. *Bit-Tech*, 7(2), 202–213. <https://doi.org/10.32877/bt.v7i2.1697>
- Juanda, R., & Yadi, I. Z. (2020). Penerapan Rule Based Dengan Algoritma Viterbi Untuk Deteksi Kesalahan Huruf Kapital Pada Karya Ilmiah. *Journal of Computer and Information Systems Ampera*, 1(1), 53–62. <https://doi.org/10.51519/journalcisa.v1i1.5>
- Juli, N., Tarigan, L. A., Surip, M., Jl, A., Iskandar, W., Baru, K., Tuan, S., & Serdang, K. D. (2024). Analisis Kesalahan Penulisan Bahasa Indonesia Pada Skripsi Mahasiswa S1 Nini Suryani Switri Ramadhani Siregar Universitas Negeri Medan beberapa aspek penting , yaitu pemakaian huruf , penulisan kata , dan pemakaian tanda baca . 2(3).
- Kurniadi, D., Farisa, S., Haviana, C., & Novianto, A. (2020). Implementasi Algoritma Cosine Similarity pada sistem arsip dokumen di Universitas Islam Sultan

- Agung. *TRANSFORMTIKA*, 17(2), 124–132.
- Nilawati, Husaini, & Salat, J. (2024). Penggunaan Metode Cosine Similarity Dan Tf-Idf Untuk Klasifikasi Judul Seminar Proposal Pada Fakultas Teknik Universitas Jabal Ghafur. *Sagita Academia Journal*, 2(1), 72–79. <https://doi.org/10.61579/sagita.v2i1.60>
- Patmawati, D. (2016). *Pedoman Penulisan Skripsi Polewali*. (59), 0–49.
- Pratama, R. P., Faisal, M., & Hanani, A. (2019). Deteksi Plagiarisme pada Dokumen Jurnal Menggunakan Metode Cosine Similarity. *SMARTICS Journal*, 5(1), 22–26. <https://doi.org/10.21067/smartics.v5i1.2848>
- Riyani, A., Zidny Naf'an #2, M., & Burhanuddin, A. (2019). Penerapan Cosine Similarity dan Pembobotan TF-IDF untuk Mendeteksi Kemiripan Dokumen. In *JLK* (Vol. 2, Number 1).
- Supiyanto, & Sriyono. (2023). Metode Cosine Similarity Untuk Mendeteksi Kemiripan Pada Dokumen Teks. *SAINS Jurnal MIPA Dan Pengajarannya*, 1(1), 1–7. <https://ejournal.uncen.ac.id/index.php/SAINS>