

**DESIGN AND CONSTRUCTION OF A WEBSITE-BASED DIGITAL
DOCUMENT CHECKING APPLICATION FOR STUDENT ASSIGNMENTS
RANCANG BANGUN APLIKASI PENGECEKAN DOKUMEN DIGITAL
TUGAS MAHASISWA BERBASIS WEBSITE**

Haeruddin^{1*}, Gunawan Putra Wahdana², Dhimas Tribuana³, Dayanti⁴, Nur Inda⁵
Universitas Prof. DR. H. M. Arifin Sallatang^{1,2}, Akademi Sekretari Manajemen Publik²,
Universitas Patria Artha³, Institut Teknologi dan Bisnis Muhammadiyah Polewali
Mandar⁵

ayankkamasee@gmail.com¹, me.gunawanputra@gmail.com², d.tribuana@gmail.com³,
dayanti.fattah@gmail.com⁴, nurinda@itbmpolman.ac.id⁵

ABSTRACT

The development of information technology has encouraged the submission of student assignments to shift from physical documents to digital documents. However, this convenience also creates new challenges, such as unstructured assignment archiving and the increasing potential for document similarity among students. This study aims to design and develop a website-based digital document checking application for student assignments that can manage documents and automatically display similarity percentages between files. The system was developed using the Waterfall method through requirement analysis, system design, implementation, testing, and evaluation stages. The technologies used include HTML, CSS, PHP, JavaScript, MySQL, XAMPP, and Sublime Text. The document checking process consists of document upload, text extraction, text preprocessing, word weighting using TF-IDF, and similarity measurement using the cosine similarity method. The data used consisted of 165 student assignment files, comprising 79 dataset files and 86 test files. Testing on five sample documents showed that the system was able to display various similarity percentages of 5%, 12%, 45%, 88%, and 92%, categorized as original, plagiarism indication, and high plagiarism. These results indicate that the application can assist lecturers in conducting an initial examination of similarity in student assignment documents. Therefore, the developed application has the potential to support more effective academic document management and improve academic integrity in higher education environments.

Keywords: digital document, student assignment, website, TF-IDF, cosine similarity, plagiarism.

ABSTRAK

Perkembangan teknologi informasi mendorong proses pengumpulan tugas mahasiswa beralih dari dokumen fisik ke dokumen digital. Namun, kemudahan tersebut menimbulkan tantangan baru, seperti pengelolaan arsip tugas yang kurang terstruktur dan meningkatnya potensi kemiripan dokumen antar mahasiswa. Penelitian ini bertujuan untuk merancang dan membangun aplikasi pengecekan dokumen digital tugas mahasiswa berbasis website yang mampu mengelola dokumen serta menampilkan persentase kemiripan antarfile secara otomatis. Sistem dikembangkan menggunakan metode Waterfall melalui tahapan analisis kebutuhan, perancangan, implementasi, pengujian, dan evaluasi. Teknologi yang digunakan meliputi HTML, CSS, PHP, JavaScript, MySQL, XAMPP, dan Sublime Text. Proses pengecekan dokumen dilakukan melalui tahapan unggah dokumen, ekstraksi teks, text preprocessing, pembobotan kata menggunakan TF-IDF, serta pengukuran kemiripan menggunakan metode cosine similarity. Data yang digunakan terdiri atas 165 file tugas mahasiswa, yaitu 79 file dataset dan 86 data uji. Hasil pengujian terhadap lima sampel dokumen menunjukkan bahwa sistem mampu menampilkan variasi persentase kemiripan sebesar 5%, 12%, 45%, 88%, dan 92%, dengan kategori orisinal, indikasi plagiat, dan plagiat tinggi. Hasil tersebut menunjukkan bahwa aplikasi dapat membantu dosen dalam melakukan pemeriksaan awal terhadap kemiripan dokumen tugas mahasiswa. Dengan demikian, aplikasi ini berpotensi mendukung pengelolaan dokumen akademik secara lebih efektif serta meningkatkan integritas akademik di lingkungan perguruan tinggi.

Kata Kunci: dokumen digital, tugas mahasiswa, website, TF-IDF, cosine similarity, plagiarisme.

*This is an open access article distributed under the terms of the Creative Commons
Attribution 4.0 International License (CC BY 4.0).*

Artikel ini adalah artikel akses terbuka yang didistribusikan di bawah ketentuan
Lisensi Creative Commons Attribution 4.0 International (CC BY 4.0).



PENDAHULUAN

Perkembangan teknologi informasi telah mendorong perubahan signifikan dalam proses pembelajaran di perguruan tinggi. Aktivitas akademik yang sebelumnya banyak dilakukan secara konvensional kini semakin bergeser ke dalam bentuk digital, termasuk proses pencarian referensi, penyusunan tugas, pengumpulan dokumen, serta evaluasi hasil belajar mahasiswa. Kemudahan akses terhadap sumber informasi digital memberikan manfaat besar bagi mahasiswa karena memungkinkan mereka memperoleh bahan belajar dari berbagai sumber secara cepat dan fleksibel. Namun, di sisi lain, kemudahan tersebut juga menimbulkan tantangan baru dalam menjaga integritas akademik, terutama terkait meningkatnya potensi kemiripan dokumen dan plagiarisme dalam tugas mahasiswa.

Plagiarisme merupakan salah satu bentuk pelanggaran akademik yang dapat mengganggu kualitas pembelajaran dan menurunkan kepercayaan terhadap hasil karya ilmiah mahasiswa. Dalam konteks pendidikan tinggi, plagiarisme tidak hanya terjadi karena kesengajaan menyalin karya orang lain, tetapi juga dapat disebabkan oleh rendahnya pemahaman mahasiswa mengenai etika penulisan, keterampilan parafrasa, dan penggunaan sitasi yang benar. Gregory dan Leeman (2021) menegaskan bahwa persepsi terhadap plagiarisme dalam lingkungan akademik sering kali dipengaruhi oleh konteks, intensi, serta pemahaman individu terhadap batas antara penggunaan sumber secara sah dan pelanggaran akademik. Oleh karena itu, pencegahan plagiarisme tidak cukup hanya dilakukan melalui pemberian sanksi, tetapi juga memerlukan dukungan sistem yang mampu membantu proses identifikasi kemiripan dokumen secara objektif.

Dalam beberapa tahun terakhir, isu plagiarisme akademik semakin kompleks seiring dengan meningkatnya penggunaan teknologi digital dan kecerdasan buatan dalam proses penyusunan tugas. Balalle dan Pannilage (2025) menjelaskan bahwa kecerdasan buatan memiliki dua sisi dalam konteks integritas akademik, yaitu dapat membantu proses pembelajaran, tetapi juga berpotensi memfasilitasi pelanggaran akademik apabila tidak digunakan secara etis. Hal serupa juga dikemukakan oleh Bittle dan El-Gayar (2025), yang menunjukkan bahwa perkembangan *Generative Artificial Intelligence* mendorong perguruan tinggi untuk meninjau ulang strategi deteksi plagiarisme, kebijakan akademik, serta mekanisme penjaminan integritas tugas mahasiswa. Dengan demikian, institusi pendidikan membutuhkan sistem yang tidak hanya mampu menyimpan dokumen secara digital, tetapi juga membantu mendeteksi tingkat kemiripan dokumen sebagai langkah awal dalam menjaga keaslian tugas mahasiswa.

Pengelolaan dokumen tugas mahasiswa dalam bentuk digital juga memerlukan sistem yang efektif. Tanpa sistem yang terstruktur, dosen dapat mengalami kesulitan dalam mengarsipkan dokumen, mencari kembali file tugas, serta membandingkan kemiripan antar dokumen mahasiswa. Pemeriksaan kemiripan secara manual membutuhkan waktu yang lama, terutama ketika jumlah dokumen yang diperiksa cukup banyak. Hal ini menjadi alasan penting untuk mengembangkan aplikasi berbasis website yang dapat membantu proses unggah, penyimpanan, pengelolaan, dan pengecekan dokumen tugas mahasiswa secara lebih efisien.

Salah satu pendekatan yang banyak digunakan dalam pengukuran kemiripan dokumen adalah kombinasi metode *Term Frequency-Inverse Document Frequency* (TF-IDF) dan *cosine similarity*. TF-IDF digunakan untuk memberikan bobot pada setiap kata berdasarkan frekuensi kemunculannya dalam suatu dokumen dan tingkat keunikannya pada kumpulan dokumen. Sementara itu, *cosine similarity* digunakan untuk menghitung tingkat kemiripan antar dokumen berdasarkan representasi vektor teks. Halim dan Lasut (2024) mengembangkan aplikasi deteksi plagiarisme dokumen berbasis website menggunakan metode TF-IDF dan *cosine similarity*. Hasil penelitian tersebut menunjukkan bahwa proses sistem mencakup *preprocessing*, perhitungan TF-IDF, dan perhitungan *cosine similarity*, dengan hasil uji yang konsisten antara perhitungan manual dan aplikasi. Temuan ini menunjukkan bahwa TF-IDF dan *cosine similarity* relevan diterapkan dalam sistem

pengecekan kemiripan dokumen akademik berbasis website.

Selain TF-IDF dan *cosine similarity*, penelitian lain juga menunjukkan bahwa pendekatan berbasis *string matching* dapat digunakan untuk mendeteksi kemiripan dokumen. Zachrias dan Gunawan (2025) mengembangkan sistem klasifikasi kemiripan dokumen menggunakan algoritma Winnowing dengan pendekatan Jaccard Coefficient. Penelitian tersebut menunjukkan bahwa sistem mampu memberikan hasil yang relevan pada dokumen berbahasa Indonesia dan dapat digunakan untuk mendukung deteksi plagiarisme akademik. Namun, pendekatan berbasis pencocokan string memiliki keterbatasan ketika berhadapan dengan dokumen yang telah diparafrasa atau dimodifikasi secara semantik. Oleh karena itu, penelitian mengenai aplikasi pengecekan dokumen perlu terus dikembangkan dengan memperhatikan kebutuhan pengguna, jenis dokumen, serta metode pengukuran kemiripan yang digunakan.

Dalam konteks akademik, aplikasi pengecekan dokumen tidak seharusnya diposisikan sebagai alat yang secara mutlak menentukan plagiarisme, melainkan sebagai alat bantu untuk memberikan indikator awal tingkat kemiripan dokumen. Pudasaini et al. (2025) menjelaskan bahwa perkembangan *Large Language Models* membuat deteksi plagiarisme semakin menantang karena teks yang dihasilkan atau dimodifikasi dapat menyerupai tulisan manusia. Oleh sebab itu, sistem deteksi perlu didukung oleh interpretasi akademik, kebijakan institusi, serta verifikasi dosen terhadap konteks isi dokumen. Dengan demikian, hasil persentase kemiripan dokumen perlu dipahami sebagai bahan pertimbangan awal, bukan sebagai satu-satunya dasar dalam menetapkan pelanggaran akademik.

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk merancang dan membangun aplikasi pengecekan dokumen digital tugas mahasiswa berbasis website. Aplikasi ini dikembangkan untuk membantu dosen dalam mengelola dokumen tugas, melakukan ekstraksi teks, memproses isi dokumen, dan menampilkan persentase kemiripan antar dokumen. Sistem dirancang menggunakan metode pengembangan perangkat lunak Waterfall dengan teknologi HTML, CSS, PHP, JavaScript, MySQL, XAMPP, dan Sublime Text. Data yang digunakan berupa dokumen tugas mahasiswa dalam bentuk digital, dengan jumlah 165 file yang terdiri atas 79 file dataset dan 86 data uji.

Kontribusi penelitian ini terletak pada pengembangan aplikasi berbasis website yang dapat digunakan sebagai alat bantu pemeriksaan awal terhadap tingkat kemiripan dokumen tugas mahasiswa. Aplikasi ini diharapkan dapat meningkatkan efektivitas pengelolaan dokumen akademik, membantu dosen dalam proses verifikasi tugas mahasiswa, serta mendukung upaya peningkatan integritas akademik di lingkungan perguruan tinggi. Selain itu, penelitian ini juga memberikan kontribusi praktis dalam penerapan metode berbasis pemrosesan teks untuk mendukung sistem pengecekan dokumen digital secara sederhana, terstruktur, dan dapat digunakan dalam konteks pembelajaran.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan *research and development* dengan model pengembangan perangkat lunak *Waterfall*. Model *Waterfall* dipilih karena memiliki tahapan pengembangan yang sistematis, berurutan, dan sesuai untuk membangun aplikasi dengan kebutuhan sistem yang telah didefinisikan sejak awal. Dalam konteks pengembangan aplikasi berbasis website, model *Waterfall* dapat digunakan untuk mengarahkan proses kerja mulai dari analisis kebutuhan, perancangan sistem, implementasi, pengujian, hingga evaluasi hasil sistem. Penggunaan model *Waterfall* masih relevan dalam pengembangan sistem informasi karena memberikan alur kerja yang jelas dan terdokumentasi pada setiap tahap pengembangan (Putri et al., 2026).

Tahapan penelitian ini terdiri atas lima tahap utama, yaitu analisis kebutuhan, perancangan sistem, implementasi, pengujian, dan evaluasi. Pada tahap analisis kebutuhan, peneliti mengidentifikasi kebutuhan fungsional dan nonfungsional sistem. Kebutuhan fungsional meliputi kemampuan sistem untuk mengunggah dokumen tugas mahasiswa,

menyimpan data dokumen, mengekstraksi teks, melakukan *text preprocessing*, menghitung bobot kata, membandingkan kemiripan antar dokumen, serta menampilkan hasil persentase kemiripan. Sementara itu, kebutuhan nonfungsional meliputi kemudahan penggunaan, kecepatan pemrosesan, kompatibilitas sistem dengan browser, dan kemampuan penyimpanan data melalui basis data.

Tahap perancangan sistem dilakukan dengan menyusun alur kerja aplikasi, rancangan basis data, serta desain antarmuka pengguna. Sistem dirancang agar dosen dapat mengunggah dokumen tugas mahasiswa dalam format digital, mengisi identitas mahasiswa, dan menjalankan proses pengecekan kemiripan dokumen. Perancangan antarmuka dibuat sederhana agar pengguna dapat mengoperasikan sistem dengan mudah. Hal ini sejalan dengan prinsip pengembangan aplikasi berbasis web, yaitu sistem harus mendukung efisiensi proses, kemudahan akses, dan interaksi pengguna yang jelas.

Tahap implementasi dilakukan dengan menerjemahkan rancangan sistem ke dalam bentuk aplikasi berbasis website. Teknologi yang digunakan meliputi HTML dan CSS untuk membangun struktur dan tampilan halaman, PHP sebagai bahasa pemrograman sisi server, JavaScript untuk mendukung interaktivitas halaman, serta MySQL sebagai basis data untuk menyimpan informasi dokumen. Selain itu, pengembangan dilakukan menggunakan XAMPP sebagai server lokal dan Sublime Text sebagai editor kode. Pemilihan teknologi tersebut disesuaikan dengan kebutuhan sistem yang bersifat web-based dan memerlukan pengelolaan dokumen serta penyimpanan data secara terstruktur.

Data yang digunakan dalam penelitian ini berupa dokumen tugas mahasiswa dalam format digital. Berdasarkan data penelitian, jumlah dokumen yang digunakan adalah 165 file, yang terdiri atas 79 file dataset dan 86 file data uji. Dokumen yang digunakan merupakan tugas mahasiswa dengan mata kuliah dan topik soal yang sama agar proses perbandingan kemiripan dokumen dapat dilakukan secara lebih relevan. Penggunaan dokumen dengan konteks tugas yang sama penting dilakukan karena pengukuran kemiripan teks akan lebih bermakna apabila dokumen yang dibandingkan berada pada domain atau topik yang sejenis.

Proses pengecekan dokumen dilakukan melalui beberapa tahapan. Tahap pertama adalah unggah dokumen, yaitu proses memasukkan file tugas mahasiswa ke dalam sistem. Setelah dokumen berhasil diunggah, sistem melakukan ekstraksi teks untuk mengambil isi teks dari dokumen digital. Tahap selanjutnya adalah *text preprocessing*, yang mencakup *case folding*, *tokenizing*, *filtering*, dan *stemming*. Tahapan ini bertujuan untuk membersihkan teks, menyeragamkan bentuk kata, menghilangkan kata yang kurang relevan, serta menghasilkan representasi teks yang lebih siap untuk diproses secara komputasional. Tahapan *preprocessing* seperti *case folding*, *filtering*, *stopword removal*, dan *stemming* banyak digunakan dalam sistem berbasis TF-IDF dan *cosine similarity* karena dapat meningkatkan kualitas representasi teks sebelum proses perhitungan kemiripan dilakukan (Albab et al., 2026).

Setelah tahap *preprocessing*, sistem melakukan pembobotan kata menggunakan metode *Term Frequency-Inverse Document Frequency* atau TF-IDF. TF-IDF digunakan untuk mengukur tingkat kepentingan suatu kata dalam dokumen berdasarkan frekuensi kemunculan kata pada dokumen dan distribusinya dalam kumpulan dokumen. Pendekatan ini banyak digunakan dalam pemrosesan bahasa alami dan pengukuran kemiripan dokumen karena mampu merepresentasikan teks ke dalam bentuk vektor numerik. Widiyanto et al. (2024) menjelaskan bahwa TF-IDF dapat digunakan untuk merepresentasikan dokumen dalam ruang vektor berdimensi tinggi, menangkap karakteristik unik isi dokumen, dan mengurangi pengaruh kata-kata umum yang muncul pada banyak dokumen.

Tahap berikutnya adalah pengukuran kemiripan dokumen menggunakan metode *cosine similarity*. Metode ini digunakan untuk menghitung kedekatan antara dua dokumen berdasarkan sudut antara vektor dokumen. Semakin besar nilai *cosine similarity*, semakin tinggi tingkat kemiripan antara dua dokumen. Kombinasi TF-IDF dan *cosine similarity* telah banyak digunakan dalam pengembangan sistem deteksi kemiripan dokumen karena mampu

menghasilkan nilai kemiripan dalam bentuk skor atau persentase yang mudah diinterpretasikan. Halim dan Lasut (2024) menunjukkan bahwa aplikasi deteksi plagiarisme berbasis web dapat menggunakan TF-IDF dan *cosine similarity* untuk menghasilkan persentase kemiripan dokumen melalui proses *preprocessing*, pembobotan, dan perhitungan kemiripan.

Secara umum, alur proses pengecekan dokumen dalam sistem ini adalah sebagai berikut:

1. Pengguna mengunggah dokumen tugas mahasiswa;
2. Sistem mengekstraksi teks dari dokumen;
3. Sistem melakukan *text preprocessing*;
4. Sistem menghitung bobot kata menggunakan TF-IDF;
5. Sistem menghitung kemiripan dokumen menggunakan *cosine similarity*;
6. Sistem menampilkan persentase kemiripan dokumen;
7. Hasil pengecekan disimpan ke dalam basis data.

Pengujian sistem dilakukan untuk mengetahui apakah aplikasi yang dikembangkan dapat menjalankan fungsi utama sesuai tujuan penelitian. Pengujian dilakukan terhadap fitur unggah dokumen, penyimpanan data, ekstraksi teks, proses *preprocessing*, perhitungan kemiripan, dan tampilan hasil persentase kemiripan. Selain itu, pengujian juga dilakukan menggunakan beberapa sampel dokumen dengan variasi tingkat kemiripan yang berbeda. Dalam naskah awal, sistem diuji menggunakan lima sampel dokumen yang menghasilkan variasi kategori, yaitu orisinal, indikasi plagiat, dan plagiat tinggi.

Hasil pengujian digunakan untuk menilai kemampuan sistem dalam membantu proses pemeriksaan awal terhadap kemiripan dokumen tugas mahasiswa. Dalam penelitian ini, nilai persentase kemiripan tidak secara langsung digunakan untuk menetapkan suatu dokumen sebagai plagiarisme, tetapi sebagai indikator awal yang perlu diverifikasi lebih lanjut oleh dosen. Pendekatan ini penting karena kemiripan teks belum tentu selalu menunjukkan pelanggaran akademik, terutama apabila terdapat kutipan, istilah teknis, atau bagian standar yang memang wajar muncul dalam dokumen akademik. Studi terbaru mengenai deteksi plagiarisme juga menegaskan bahwa sistem deteksi perlu dipahami sebagai alat bantu, karena pendekatan berbasis teks masih memiliki keterbatasan dalam mendeteksi parafrasa, perubahan struktur kalimat, atau kemiripan semantik yang lebih kompleks (Sajid et al., 2025).

Dengan metode tersebut, penelitian ini menghasilkan aplikasi berbasis website yang dapat membantu dosen dalam mengelola dokumen tugas mahasiswa dan menampilkan tingkat kemiripan antar dokumen secara otomatis. Sistem yang dikembangkan tidak dimaksudkan untuk menggantikan penilaian akademik dosen, melainkan sebagai alat bantu untuk mempercepat proses identifikasi awal terhadap dokumen yang memiliki tingkat kemiripan tinggi.

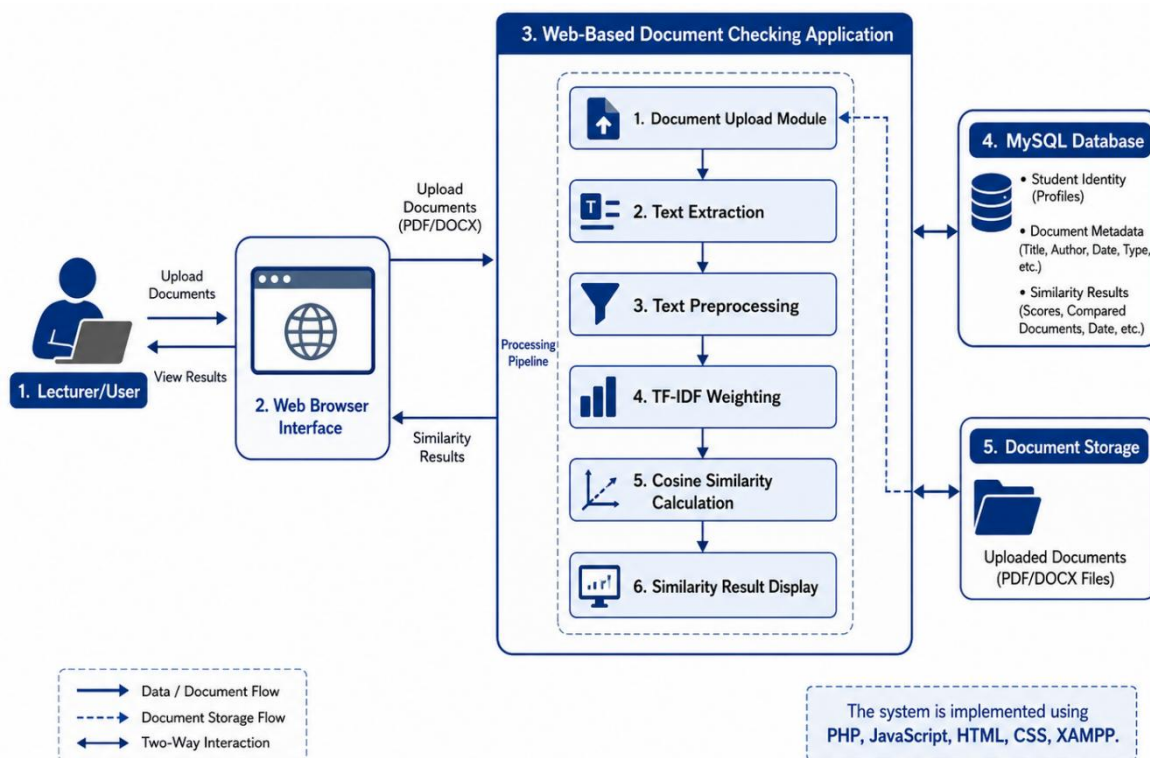
HASIL DAN PEMBAHASAN

Implementasi Sistem

Penelitian ini menghasilkan aplikasi pengecekan dokumen digital tugas mahasiswa berbasis website yang dirancang untuk membantu dosen dalam mengelola dan memeriksa tingkat kemiripan dokumen tugas mahasiswa. Sistem dibangun dengan fitur utama berupa unggah dokumen, penyimpanan data tugas, ekstraksi teks, *text preprocessing*, pembobotan kata menggunakan TF-IDF, serta pengukuran kemiripan dokumen menggunakan metode *cosine similarity*.

Aplikasi ini dikembangkan menggunakan HTML, CSS, PHP, JavaScript, dan MySQL. Sistem dijalankan melalui server lokal menggunakan XAMPP. Secara umum, sistem memiliki alur kerja yang dimulai dari proses unggah dokumen tugas mahasiswa, ekstraksi isi dokumen, pemrosesan teks, perhitungan bobot kata, perhitungan nilai kemiripan, hingga penampilan hasil persentase kemiripan dokumen.

Berdasarkan rancangan pada naskah, aplikasi ini memiliki halaman unggah dokumen yang memungkinkan dosen memasukkan nama mahasiswa, memilih file tugas, serta mengunggah dokumen ke dalam sistem. Dokumen yang telah diunggah kemudian diproses oleh sistem untuk dihitung tingkat kemiripannya dengan dokumen lain yang tersimpan dalam basis data.



Gambar 1. Alur Kerja Sistem Pengecekan Dokumen Digital

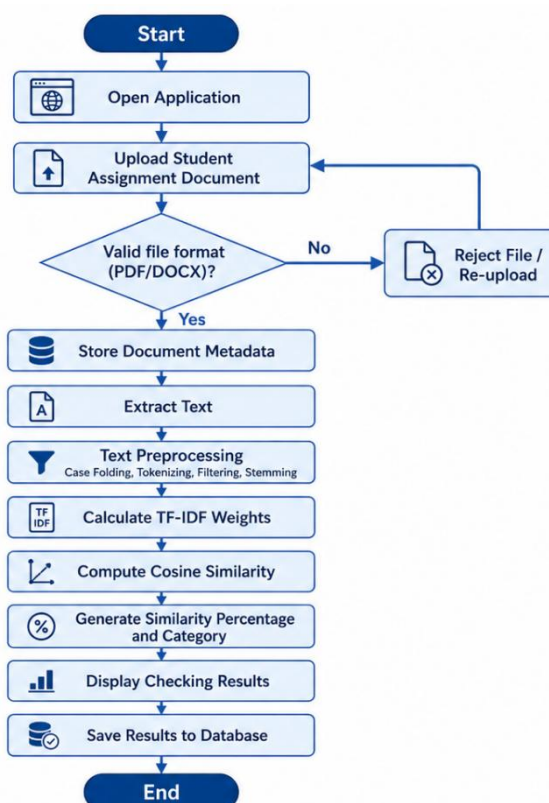
Gambar 1 menunjukkan alur kerja aplikasi pengecekan dokumen digital tugas mahasiswa. Proses dimulai dari unggah file dokumen, kemudian sistem melakukan pemeriksaan format file. Jika format file sesuai, sistem akan melakukan ekstraksi teks dari dokumen. Teks hasil ekstraksi selanjutnya diproses melalui tahapan *preprocessing*, yaitu *case folding*, *tokenizing*, *filtering*, dan *stemming*. Setelah itu, sistem melakukan pembobotan kata menggunakan TF-IDF dan menghitung nilai kemiripan antar dokumen menggunakan *cosine similarity*. Hasil akhir dari proses ini berupa persentase kemiripan dokumen yang ditampilkan kepada pengguna.

Proses Pengecekan Kemiripan Dokumen

Proses pengecekan kemiripan dokumen dilakukan melalui beberapa tahapan utama. Tahap pertama adalah unggah dokumen, yaitu dosen memasukkan file tugas mahasiswa ke dalam sistem. File yang digunakan dapat berupa dokumen digital, seperti PDF atau DOCX. Setelah dokumen berhasil diunggah, sistem melakukan ekstraksi teks untuk mengambil isi dokumen yang akan dianalisis.

Tahap berikutnya adalah *text preprocessing*. Pada tahap ini, teks yang telah diekstraksi diproses agar siap digunakan dalam perhitungan kemiripan dokumen. Tahapan *preprocessing* meliputi:

1. **Case folding**, yaitu mengubah seluruh huruf menjadi bentuk yang seragam.
2. **Tokenizing**, yaitu memecah teks menjadi unit kata.
3. **Filtering**, yaitu menghapus kata yang tidak relevan atau kata umum.
4. **Stemming**, yaitu mengubah kata ke bentuk dasarnya.



Gambar 2. Flowchart System

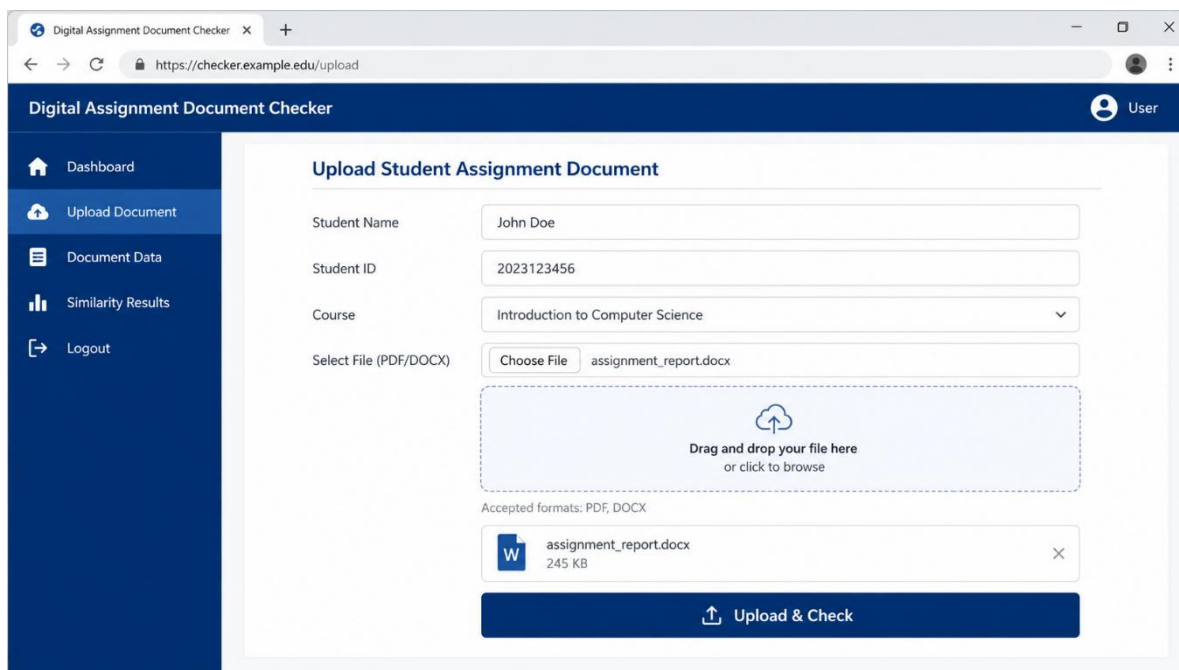
Setelah tahap *preprocessing*, sistem melakukan pembobotan kata menggunakan metode TF-IDF. Metode ini digunakan untuk memberikan bobot pada setiap kata berdasarkan frekuensi kemunculannya dalam dokumen dan tingkat keunikannya dalam kumpulan dokumen. Kata yang sering muncul dalam satu dokumen tetapi jarang muncul pada dokumen lain akan memiliki bobot lebih tinggi.

Tahap terakhir adalah perhitungan nilai kemiripan menggunakan *cosine similarity*. Metode ini menghitung kedekatan antar dokumen berdasarkan sudut antara vektor dokumen. Semakin tinggi nilai *cosine similarity*, semakin tinggi pula tingkat kemiripan antara dua dokumen. Hasil perhitungan kemudian dikonversi ke dalam bentuk persentase agar lebih mudah dipahami oleh dosen sebagai pengguna sistem.

Implementasi Antarmuka Aplikasi

Salah satu tampilan utama dalam sistem adalah halaman unggah tugas mahasiswa. Pada halaman ini, dosen dapat memasukkan identitas mahasiswa, memilih file tugas, serta mengunggah dokumen ke dalam sistem. Antarmuka dibuat sederhana agar mudah digunakan oleh dosen tanpa memerlukan keterampilan teknis yang kompleks.

Gambar 3 memperlihatkan halaman unggah tugas mahasiswa pada aplikasi yang dikembangkan. Pada halaman tersebut tersedia kolom nama mahasiswa, tombol pemilihan file tugas, serta tombol untuk mengirim data ke sistem. Setelah dosen mengunggah file, sistem akan menyimpan dokumen ke dalam basis data dan memproses isi dokumen untuk dilakukan pengecekan kemiripan. Tampilan ini dirancang sederhana agar proses pengelolaan dokumen tugas mahasiswa dapat dilakukan secara praktis dan efisien.



Gambar 3. Tampilan Halaman Unggah Tugas Mahasiswa

Hasil Pengujian Sistem

Pengujian sistem dilakukan menggunakan lima sampel dokumen tugas mahasiswa dengan tingkat kemiripan yang berbeda. Pengujian bertujuan untuk mengetahui kemampuan sistem dalam membaca dokumen, menghitung jumlah kata, memproses dokumen, dan menampilkan persentase kemiripan. Hasil pengujian menunjukkan bahwa sistem mampu memberikan variasi nilai kemiripan dokumen, mulai dari kategori orisinal, indikasi plagiat, hingga plagiat tinggi.

File Name	Word Count	Processing Time	Similarity	Category
Tugas_AI_01.docx	450	1.2 s	12%	Original
Tugas_AI_02.docx	512	1.5 s	88%	High Plagiarism
Analisis_Sistem.pdf	1200	3.1 s	45%	Plagiarism Indication
Resume_TI.docx	320	0.9 s	5%	Original
Tugas_Final_V2.pdf	850	2.4 s	92%	High Plagiarism

(i) Similarity results are used as an initial indicator and require lecturer verification.

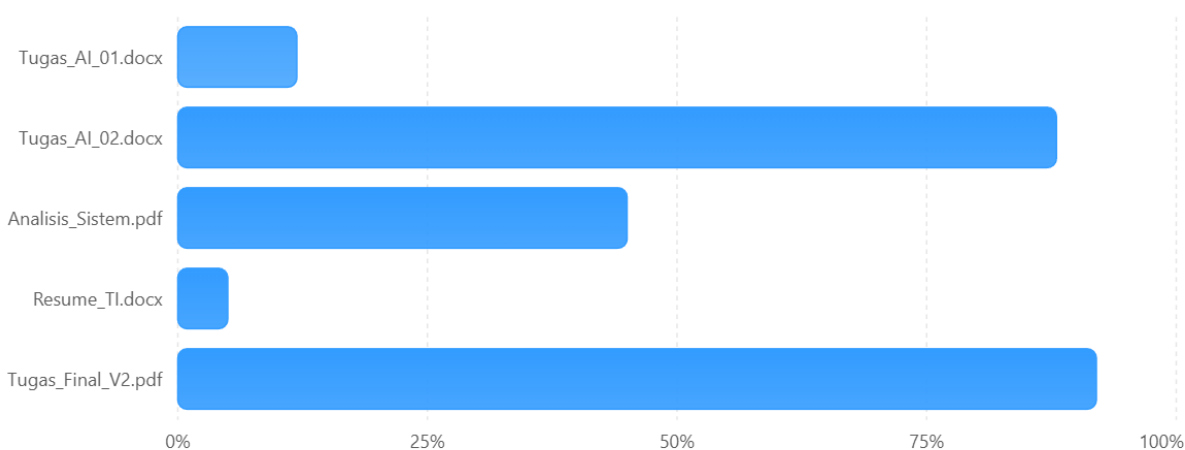
Gambar 4. Hasil Pengujian Kemiripan Dokumen

Tabel 1. Hasil Pengujian Kemiripan Dokumen

No	Nama File	Jumlah Kata	Waktu Proses	Kemiripan	Keterangan
1	Tugas_AI_01.docx	450	1,2 detik	12%	Orisinal
2	Tugas_AI_02.docx	512	1,5 detik	88%	Plagiat Tinggi
3	Analisis_Sistem.pdf	1.200	3,1 detik	45%	Indikasi Plagiat
4	Resume_TI.docx	320	0,9 detik	5%	Orisinal
5	Tugas_Final_V2.pdf	850	2,4 detik	92%	Plagiat Tinggi

Berdasarkan Tabel 1, sistem mampu menampilkan persentase kemiripan dokumen dengan hasil yang bervariasi. Dokumen Resume_TI.docx memiliki tingkat kemiripan paling rendah, yaitu 5%, sehingga dikategorikan sebagai orisinal. Dokumen Tugas_AI_01.docx memiliki tingkat kemiripan 12% dan juga termasuk kategori orisinal. Sementara itu, dokumen Analisis_Sistem.pdf memiliki tingkat kemiripan 45%, sehingga dikategorikan sebagai indikasi plagiat dan memerlukan pemeriksaan lebih lanjut oleh dosen.

Dua dokumen lainnya, yaitu Tugas_AI_02.docx dan Tugas_Final_V2.pdf, memiliki tingkat kemiripan yang sangat tinggi, masing-masing sebesar 88% dan 92%. Nilai tersebut menunjukkan bahwa kedua dokumen memiliki kemiripan teks yang tinggi terhadap dokumen lain dalam basis data, sehingga dikategorikan sebagai plagiat tinggi.

**Gambar 5.** Visualisasi Persentase Kemiripan Dokumen

Gambar 5 menunjukkan visualisasi persentase kemiripan lima sampel dokumen tugas mahasiswa. Grafik tersebut memperlihatkan bahwa dua dokumen memiliki tingkat kemiripan sangat tinggi, yaitu Tugas_Final_V2.pdf sebesar 92% dan Tugas_AI_02.docx sebesar 88%. Sementara itu, dokumen Resume_TI.docx dan Tugas_AI_01.docx memiliki tingkat kemiripan rendah. Visualisasi ini memudahkan dosen dalam mengidentifikasi dokumen yang perlu mendapat perhatian lebih lanjut.

Waktu Proses Sistem

Selain menampilkan persentase kemiripan, sistem juga menunjukkan waktu proses untuk setiap dokumen. Berdasarkan hasil pengujian, waktu proses berkisar antara 0,9 detik hingga 3,1 detik. Dokumen dengan jumlah kata lebih banyak cenderung membutuhkan waktu proses yang lebih lama. Misalnya, dokumen Analisis_Sistem.pdf memiliki 1.200 kata dan membutuhkan waktu proses 3,1 detik. Sementara itu, dokumen Resume_TI.docx yang memiliki 320 kata hanya membutuhkan waktu proses 0,9 detik.

Hasil ini menunjukkan bahwa jumlah kata dalam dokumen berpengaruh terhadap waktu pemrosesan. Semakin banyak kata yang harus diekstraksi dan dihitung bobotnya, semakin besar pula waktu yang dibutuhkan sistem untuk menghasilkan nilai kemiripan. Meskipun demikian, waktu proses yang ditampilkan masih berada dalam kategori cepat untuk

kebutuhan pemeriksaan awal dokumen tugas mahasiswa.

Tabel 2. Hubungan Jumlah Kata dan Waktu Proses

Nama File	Jumlah Kata	Waktu Proses
Resume_TI.docx	320	0,9 detik
Tugas_AI_01.docx	450	1,2 detik
Tugas_AI_02.docx	512	1,5 detik
Tugas_Final_V2.pdf	850	2,4 detik
Analisis_Sistem.pdf	1.200	3,1 detik

Berdasarkan Tabel 2, terdapat kecenderungan bahwa dokumen dengan jumlah kata lebih besar membutuhkan waktu proses lebih lama. Hal ini wajar karena sistem harus melakukan lebih banyak proses, mulai dari ekstraksi teks, pemecahan kata, penghapusan kata tidak relevan, pembobotan TF-IDF, hingga perhitungan *cosine similarity*.

Hasil penelitian menunjukkan bahwa aplikasi pengecekan dokumen digital tugas mahasiswa berbasis website mampu menjalankan fungsi utama yang dibutuhkan dalam pengelolaan dan pemeriksaan dokumen. Sistem mampu menerima unggahan dokumen, mengekstraksi teks, melakukan pemrosesan kata, menghitung bobot kata, serta menampilkan persentase kemiripan dokumen. Dengan demikian, aplikasi ini dapat menjadi alat bantu bagi dosen dalam melakukan pemeriksaan awal terhadap potensi kemiripan tugas mahasiswa.

Penerapan metode TF-IDF dan *cosine similarity* terbukti sesuai untuk kebutuhan sistem pengecekan dokumen berbasis teks. TF-IDF membantu sistem dalam membedakan kata yang memiliki bobot penting dengan kata umum yang sering muncul dalam banyak dokumen. Hal ini penting karena dokumen akademik sering memiliki kata-kata umum yang sama, seperti istilah teknis, kata penghubung, atau frasa akademik. Dengan pembobotan TF-IDF, sistem dapat lebih fokus pada kata yang memiliki kontribusi penting dalam membedakan isi dokumen.

Sementara itu, *cosine similarity* membantu sistem dalam menghitung kedekatan antar dokumen berdasarkan representasi vektor teks. Hasil perhitungan dalam bentuk persentase memudahkan dosen untuk memahami tingkat kemiripan dokumen. Persentase kemiripan rendah dapat mengindikasikan bahwa dokumen relatif orisinal, sedangkan persentase kemiripan tinggi dapat menjadi tanda bahwa dokumen perlu diperiksa lebih lanjut.

Meskipun demikian, hasil persentase kemiripan tidak dapat langsung dijadikan dasar tunggal untuk menyatakan bahwa suatu dokumen melakukan plagiarisme. Kemiripan teks dapat terjadi karena penggunaan istilah teknis yang sama, kesamaan format tugas, kutipan yang benar, atau instruksi soal yang seragam. Oleh karena itu, sistem ini lebih tepat diposisikan sebagai alat bantu deteksi awal, bukan sebagai sistem penentu akhir pelanggaran akademik.

Hasil pengujian menunjukkan bahwa dokumen dengan tingkat kemiripan tinggi, seperti Tugas_AI_02.docx dan Tugas_Final_V2.pdf, dapat langsung ditandai sebagai dokumen yang perlu diperiksa lebih mendalam oleh dosen. Hal ini membuat proses pemeriksaan tugas menjadi lebih efisien karena dosen dapat memprioritaskan dokumen dengan persentase kemiripan tinggi. Sebaliknya, dokumen dengan tingkat kemiripan rendah dapat diproses lebih cepat karena memiliki risiko kemiripan yang relatif kecil.

Dari sisi efektivitas sistem, aplikasi ini memberikan manfaat dalam dua aspek utama. Pertama, sistem membantu pengelolaan dokumen tugas mahasiswa secara lebih terstruktur. Dokumen yang sebelumnya tersebar dalam berbagai media penyimpanan dapat disimpan dan dikelola melalui sistem berbasis website. Kedua, sistem membantu proses pemeriksaan kemiripan dokumen secara otomatis, sehingga dapat mengurangi beban kerja dosen dalam melakukan pemeriksaan manual.

Namun, sistem yang dikembangkan masih memiliki beberapa keterbatasan. Pertama, kemampuan sistem sangat bergantung pada kualitas ekstraksi teks dari dokumen. Jika

dokumen berupa hasil pindai atau gambar, sistem memerlukan dukungan OCR agar teks dapat dibaca. Kedua, sistem hanya membandingkan dokumen dengan file yang tersedia dalam basis data, sehingga belum dapat memeriksa kemiripan dengan sumber eksternal dari internet. Ketiga, metode TF-IDF dan *cosine similarity* lebih kuat dalam mendeteksi kesamaan tekstual, tetapi masih terbatas dalam mendeteksi parafrasa atau perubahan kalimat yang memiliki makna sama.

Berdasarkan keterbatasan tersebut, pengembangan selanjutnya dapat diarahkan pada integrasi dengan teknologi OCR, perluasan basis data dokumen, serta penerapan pendekatan semantik berbasis *natural language processing* atau *machine learning*. Dengan demikian, sistem tidak hanya mampu mendeteksi kemiripan kata, tetapi juga dapat menganalisis kedekatan makna antar dokumen.

Hasil penelitian ini memiliki implikasi praktis bagi perguruan tinggi, khususnya dalam pengelolaan tugas mahasiswa dan peningkatan integritas akademik. Aplikasi yang dikembangkan dapat digunakan sebagai sistem pendukung bagi dosen untuk melakukan pemeriksaan awal terhadap dokumen tugas mahasiswa. Dengan adanya sistem ini, proses pengecekan tidak lagi sepenuhnya dilakukan secara manual, sehingga waktu pemeriksaan dapat menjadi lebih efisien.

Selain itu, aplikasi ini juga dapat menjadi media edukasi bagi mahasiswa agar lebih berhati-hati dalam menyusun tugas akademik. Mahasiswa dapat memahami bahwa dokumen digital yang dikumpulkan memiliki kemungkinan untuk diperiksa tingkat kemiripannya. Hal ini dapat mendorong mahasiswa untuk lebih memperhatikan orisinalitas tulisan, penggunaan sitasi, dan etika akademik.

Dari sisi pengembangan sistem, penelitian ini menunjukkan bahwa teknologi web sederhana dapat digunakan untuk membangun aplikasi pengecekan dokumen yang fungsional. Dengan memanfaatkan PHP, MySQL, TF-IDF, dan *cosine similarity*, sistem dapat memberikan hasil yang cukup informatif untuk mendukung proses akademik di perguruan tinggi.

KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, aplikasi pengecekan dokumen digital tugas mahasiswa berbasis website berhasil dirancang dan dibangun untuk membantu proses pengelolaan serta pemeriksaan tingkat kemiripan dokumen tugas mahasiswa. Sistem dikembangkan menggunakan metode Waterfall melalui tahapan analisis kebutuhan, perancangan, implementasi, pengujian, dan evaluasi. Fitur utama yang dihasilkan meliputi unggah dokumen, penyimpanan data tugas mahasiswa, ekstraksi teks, *text preprocessing*, pembobotan kata menggunakan TF-IDF, serta pengukuran kemiripan dokumen menggunakan metode *cosine similarity*.

Hasil pengujian menunjukkan bahwa sistem mampu memproses dokumen tugas mahasiswa dan menampilkan persentase kemiripan dalam beberapa kategori, yaitu orisinal, indikasi plagiat, dan plagiat tinggi. Pada pengujian terhadap lima sampel dokumen, sistem menghasilkan tingkat kemiripan yang bervariasi, yaitu 5%, 12%, 45%, 88%, dan 92%. Dokumen dengan tingkat kemiripan rendah dikategorikan sebagai orisinal, sedangkan dokumen dengan tingkat kemiripan tinggi perlu mendapatkan pemeriksaan lanjutan oleh dosen. Hasil ini menunjukkan bahwa aplikasi dapat digunakan sebagai alat bantu awal dalam mengidentifikasi dokumen yang memiliki potensi kemiripan tinggi.

Aplikasi yang dikembangkan juga dapat membantu dosen dalam mengelola dokumen tugas mahasiswa secara lebih terstruktur. Melalui sistem berbasis website, proses pengunggahan, penyimpanan, dan pengecekan dokumen dapat dilakukan secara lebih praktis dibandingkan pemeriksaan manual. Dengan demikian, sistem ini berpotensi meningkatkan efektivitas dan efisiensi proses pemeriksaan tugas mahasiswa, khususnya pada kelas dengan jumlah dokumen yang cukup banyak.

Meskipun demikian, hasil persentase kemiripan yang ditampilkan oleh sistem tidak dapat

dijadikan satu-satunya dasar untuk menetapkan plagiarisme. Hasil tersebut perlu diposisikan sebagai indikator awal yang tetap memerlukan verifikasi akademik oleh dosen, terutama untuk membedakan antara kemiripan yang disebabkan oleh plagiarisme, penggunaan istilah teknis yang sama, format tugas yang seragam, atau kutipan yang sah. Selain itu, sistem masih memiliki keterbatasan karena hanya membandingkan dokumen dengan data yang tersimpan dalam basis data dan belum mencakup sumber eksternal dari internet.

Secara keseluruhan, penelitian ini menunjukkan bahwa penerapan TF-IDF dan *cosine similarity* pada aplikasi berbasis website dapat mendukung proses pengecekan kemiripan dokumen digital tugas mahasiswa. Untuk pengembangan selanjutnya, sistem dapat ditingkatkan dengan menambahkan teknologi OCR untuk membaca dokumen hasil pindai, integrasi pencarian sumber eksternal, serta pendekatan berbasis *semantic similarity* atau *machine learning* agar sistem mampu mendeteksi kemiripan makna dan parafrasa secara lebih baik.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada pihak-pihak yang mendukung pelaksanaan penelitian ini.

DAFTAR PUSTAKA

- Balalle, H., & Pannilage, S. (2025). Reassessing academic integrity in the age of AI: A systematic literature review on AI and academic integrity. *Social Sciences & Humanities Open*, 11, 101299. <https://doi.org/10.1016/j.ssaho.2025.101299>
- Bittle, K., & El-Gayar, O. (2025). Generative AI and academic integrity in higher education: A systematic review and research agenda. *Information*, 16(4), 296. <https://doi.org/10.3390/info16040296>
- Foltýnek, T., Meuschke, N., & Gipp, B. (2020). Academic plagiarism detection: A systematic literature review. *ACM Computing Surveys*, 52(6), 1–42. <https://doi.org/10.1145/3345317>
- Gregory, A., & Leeman, J. (2021). *On the perception of plagiarism in academia: Context and intent*. arXiv. <https://arxiv.org/abs/2104.00574>
- 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>
- Kaur, N. (2024). A review on string-based text similarity techniques in computational analysis. *International Journal of Intelligent Systems and Applications in Engineering*. <https://ijisae.org/index.php/IJISAE/article/view/7629>
- Meuschke, N., Stange, V., Schubotz, M., Kramer, M., & Gipp, B. (2019). Improving academic plagiarism detection for STEM documents by analyzing mathematical content and citations. In *Proceedings of the 19th ACM/IEEE Joint Conference on Digital Libraries (JCDL)* (pp. 120–129). IEEE. <https://doi.org/10.1109/JCDL.2019.00026>
- Navaro, K. F., Wardana, S. R., & Hapsari, R. K. (2023). Journal article plagiarism detection using Latent Semantic Analysis (LSA). *Prosiding Seminar Nasional Sains dan Teknologi Terapan*. <https://ejurnal.itats.ac.id/sntekpan/article/view/5204>
- Pudasaini, S., Miralles-Pechuán, L., Lillis, D., & Llorens Salvador, M. (2025). Survey on AI-generated plagiarism detection: The impact of large language models on academic integrity.

Journal of Academic Ethics, 23, 1137–1170. <https://doi.org/10.1007/s10805-024-09576-x>

Sadhin, I. H., Hassan, T., & Nayim, M. A. M. (2024). Plagiarism detection using artificial intelligence. *International Journal of Computer and Information System*, 5(2), 102–108. <https://doi.org/10.29040/ijcis.v5i2.170>

Virginia, C., & Alamsyah, D. (2026). Plagiarism detection in English academic documents using a lexical semantic hybrid and support vector machine. *INOVTEK Polbeng – Seri Informatika*, 11(1), 96–107. <https://doi.org/10.35314/2zz12581>

Zachrias, U., & Gunawan, W. (2025). Classification of document similarity using Winnowing algorithm with Jaccard Coefficient approach. *Science of Information & Technology Applied*. <https://ejournal.bacadulu.net/index.php/sinta/article/view/87>