

Sistem Deteksi Plagiarisme Tugas Akhir Mahasiswa Berbasis Natural Language Processing Menggunakan Algoritma Jaro-Winkler dan TF-IDF

Darmanto^{*1}, Novi Indah Pradasari², Eka Wahyudi³

^{1,2,3}Program Studi Teknologi Informasi, Jurusan Teknik Elektro dan Teknik Informatika, Politeknik Negeri Ketapang

Email: ^{*1}darmanto@politap.ac.id, ²novi.ip@politap.ac.id, ³eka.wahyudi@politap.ac.id

(Naskah masuk: 27 Desember 2023, diterima untuk diterbitkan: 27 Desember 2023)

Abstrak: Penelitian ini bertujuan untuk mengatasi permasalahan plagiarisme dan kesalahan penulisan yang sering terjadi pada karya tulis mahasiswa di lingkungan akademik, khususnya di Politeknik Negeri Ketapang (Politap). Saat ini, pengajuan judul tugas akhir masih dilakukan secara konvensional yang menyebabkan sulitnya pengelolaan data administrasi tugas akhir dan rawan terdapat risiko plagiarisme. Oleh karena itu, penelitian ini di fokuskan dalam mengembangkan sistem deteksi plagiarisme tugas akhir mahasiswa berbasis Natural Language Processing (NLP) dengan menggunakan algoritma Jaro-Winkler dan TF-IDF. Penelitian ini menggunakan metode studi literatur dan observasi untuk pengumpulan data dan kebutuhan sistem, sedangkan untuk pengembangan sistem menggunakan Agile Software Development model Scrum. Data yang digunakan dalam penelitian ini merupakan data karya tugas akhir mahasiswa yang terdiri dari 20 dokumen Tugas Akhir Mahasiswa Jurusan Teknik Elektro dan Teknik Informatika Tahun 2022 diambil secara acak untuk digunakan sebagai Dataset untuk menguji efektifitas kinerja sistem yang dibangun. Hasil dari penelitian ini adalah sebuah aplikasi pendeteksi plagiarisme dan kesalahan penulisan yang dapat membantu mahasiswa dan dosen dalam mencegah terjadinya kasus plagiarisme dan kesalahan penulisan pada karya tugas akhir dengan mengutamakan kejujuran dan integritas akademik.

Kata Kunci – Text Mining; Natural Language Processing; Jaro-Winkler; Plagiarism

Student Final Project Plagiarism Detection System Based on Natural Language Processing Using the Jaro-Winkler and TF-IDF Algorithms

Abstract: This research aims to overcome the problem of plagiarism and writing errors that often occur in students' written work in the academic environment, especially at the Ketapang State Polytechnic (Politap). Currently, submission of final assignment titles is still done conventionally, which makes it difficult to manage final assignment administrative data and is prone to the risk of plagiarism. Therefore, this research focuses on developing a student final assignment plagiarism detection system based on Natural Language Processing (NLP) using the Jaro-Winkler and TF-IDF algorithms. This research uses literature study and observation methods to collect data and system requirements, while for system development it uses the Agile Software Development Scrum model. The data used in this research is student final assignment data which consists of 20 Final Assignment documents for students majoring in Electrical Engineering and Information Engineering in 2022 taken randomly to be used as a dataset to test the effectiveness of the performance of the system being built. The result of this research is a plagiarism and writing error detection application that can help students and lecturers prevent cases of plagiarism and writing errors in final assignments by prioritizing honesty and academic integrity.

Keywords – Text Mining; Natural Language Processing; Jaro-Winkler; Plagiarism

1. PENDAHULUAN

Pada era digital saat ini, akses mudah ke informasi melalui internet telah memudahkan tindakan plagiarisme. Tindakan plagiarisme pada tugas akhir mahasiswa merupakan isu serius

yang memengaruhi berbagai pihak, terutama lembaga pendidikan tinggi. Dampak dari plagiarisme meluas, merugikan mahasiswa, dosen, lembaga pendidikan, dan masyarakat secara keseluruhan. Plagiarisme dapat mengurangi integritas akademik dan merusak citra institusi pendidikan, serta mengurangi kualitas lulusan.

Saat ini, mekanisme pengajuan judul tugas akhir mahasiswa di Politeknik Negeri Ketapang (Politap) masih berlangsung secara konvensional, melalui serangkaian tahapan. Mahasiswa harus memilih dosen pembimbing yang sesuai dengan topik tugas akhir mereka, merumuskan topik, dan mengajukan proposal tugas akhir. Pengelolaan administrasi tugas akhir yang sederhana dengan penggunaan excel dan word serta kurangnya integrasi dengan database, menyebabkan kesulitan dalam menentukan keaslian judul tugas akhir dan mengakibatkan kerentanan terhadap plagiarisme. Diperlukan inovasi untuk mengatasi masalah ini.

Upaya untuk mencegah dan mengatasi plagiarisme memerlukan pengembangan sistem deteksi plagiarisme yang efektif. Teknologi Natural Language Processing (NLP) telah digunakan dalam pengembangan sistem deteksi plagiarisme tugas akhir. Namun, terdapat kelemahan dalam metode dan algoritma yang digunakan. Oleh karena itu, perlu pengembangan sistem deteksi plagiarisme tugas akhir yang lebih efektif.

Penelitian deteksi plagiarisme menggunakan teknik NLP telah banyak dilakukan dalam beberapa tahun terakhir. Beberapa studi terkait telah memperkenalkan metode baru yang lebih canggih dan lebih akurat dalam mendeteksi plagiarisme, seperti Penelitian[1] evaluasi metode deteksi plagiarisme akademik yang berbeda menggunakan fuzzy multi-criteria decision-making dan memberikan rekomendasi untuk pengembangan sistem deteksi plagiarisme yang efisien. Penelitian menghasilkan cetak biru sistem untuk menggunakan algoritma seperti NLP dan Machine Learning dengan jaringan saraf buatan menjadi teknik pembelajaran mesin yang paling berdampak untuk membangun alat pemeriksaan plagiarisme.

Penelitian terkait dengan sistem pendeteksi plagiarisme Tugas Akhir oleh[2] menciptakan Plagiarism Checker menggunakan Algoritma Jaro-Winkler untuk mengidentifikasi kemiripan antara abstrak calon skripsi dengan abstrak yang telah tersimpan di basis data E-Read. Hasil pengujian memperoleh akurasi sebesar 40%, presisi sebesar 33,3%, recall sebesar 100%, dan F-Measure sebesar 35%. Sedangkan[3] Algoritma TF-IDF melalui beberapa tahap proses translasi, tokenisasi, eliminasi, stemming, dan perhitungan cosines similarity. Tujuannya untuk memperoleh hasil presentase kesamaan dokumen yang diuji. Penelitian[4] menggunakan Algoritma Rabin-Karp dengan text preprocessing seperti case folding, filtering, dan tokenizing berhasil mendeteksi similaritas judul tugas akhir dengan akurasi sebesar 70%.

Penelitian[5] menggunakan Jaro-Winkler untuk pendeteksi kesamaan kata dalam aplikasi english conversation. Algoritma ini membandingkan hasil deteksi ucapan suara dengan kata kunci, sehingga siswa dapat melatih percakapan Bahasa Inggris dengan tepat dan mengetahui nilai perbandingannya. Hasil pengujian menunjukkan bahwa aplikasi ini mencapai nilai perbandingan maksimal 100% dan minimal 68.1%. Algoritma Jaro-Winkler memiliki rata-rata similarity sebesar 80.92 %, Algoritma Levensthein Distance sebesar 49.43%[6]. Penggunaan Algoritma Jaro-Winkler sebagai autocorrect dan spelling suggestion bahasa Indonesia oleh[7] diharapkan dapat membantu News Director BMS TV untuk memeriksa dan memperbaiki kesalahan penulisan kata secara otomatis serta memberi saran penulisan ejaan kata yang benar dalam bahasa Indonesia.

Sementara penggunaan Algoritma TF-IDF[8] untuk text mining sangat membantu pengguna dalam mendapatkan informasi pada kumpulan dokumen. Selain itu[9] cosine similarity dengan pembobotan kata menggunakan metode TF-IDF dan menerapkan text preprocessing untuk mengetahui syarah hadits. Hasil uji coba dengan pengujian confusion matrix didapatkan: recall 88.7%, precision 100%, accuracy 88,73 %, dan error rate 11,27 %. Implementasi Algoritma Jaro-Winkler Distance Untuk Sistem Pendeteksi Plagiarisme Pada Dokumen Skripsi[10] dalam penelitian ini, aplikasi yang menggunakan bahasa pemrograman Java dengan NetBeans IDE 8.0 dikembangkan untuk mendeteksi plagiarisme dalam dokumen skripsi. Sistem yang dibangun dapat membandingkan dokumen asli dengan dokumen uji untuk mengidentifikasi tingkat kemiripan (similarity) di antara keduanya.

Penelitian[11] tujuan penelitian ini adalah untuk mengurangi kasus plagiarisme. Sistem yang dirancang menggunakan algoritma TF-IDF dan melalui serangkaian proses, termasuk translasi, tokenisasi, eliminasi, stemming, dan perhitungan kesamaan kosinus. Hasil pengujian menunjukkan bahwa sistem ini mampu mendeteksi tingkat kesamaan dalam dokumen yang diuji. Penelitian yang dilakukan oleh[12] mengembangkan sistem pendeteksi kemiripan naskah skripsi mahasiswa di Jurusan Teknik Informatika dan Komputer Universitas Negeri Makassar menggunakan PHP dengan Framework Laravel. Sistem ini divalidasi oleh ahli sistem, memenuhi standar kualitas ISO 25010, dan layak digunakan untuk mengecek kemiripan dokumen. Penelitian tentang deteksi plagiarisme juga dilakukan oleh[13] penelitian menganalisis persepsi dosen terhadap kebijakan penggunaan Turnitin dalam mendeteksi plagiarisme skripsi mahasiswa. Data dikumpulkan melalui wawancara dengan 7 dosen pembimbing skripsi dan dianalisis menggunakan metode tematik. Temuan penelitian mencakup persepsi dosen terhadap penggunaan Turnitin dan kebijakan penggunaannya.

Penelitian ini akan memfokuskan pada pengembangan sistem deteksi plagiarisme tugas akhir berbasis Natural Language Processing (NLP) menggunakan algoritma Jaro-Winkler dan TF-IDF. NLP merupakan teknologi yang memungkinkan komputer untuk memahami bahasa manusia dan menerjemahkannya kedalam bentuk yang dapat diproses oleh mesin. Algoritma Jaro-Winkler merupakan sebuah metode algoritma yang digunakan untuk mengukur seberapa mirip dua teks dan terdiri dari tiga komponen dasar: (1) menghitung panjang string atau kata, (2) menentukan urutan huruf pada kedua kata, dan (3) mencari transposisi. Semakin besar nilai Jaro-Winkler Distance antara dua teks, semakin tinggi tingkat kemiripannya[1]. Sedangkan Term Frequency Inverse Document Frequency (TF-IDF) adalah metode yang digunakan untuk memberikan nilai atau bobot yang menunjukkan seberapa pentingnya sebuah kata (term) dalam konteks dokumen. Ini didasarkan pada seberapa sering kata itu muncul dalam dokumen (TF) dan seberapa jarang kata itu muncul dalam koleksi seluruh dokumen (IDF)[2].

Berdasarkan permasalahan yang sudah di paparkan, maka penelitian ini bertujuan untuk menggunakan Algoritma Jaro-Winkler Distance sebagai algoritma untuk mendeteksi kesalahan penulisan dalam bahasa Indonesia pada karya akhir mahasiswa dan Algoritma TF-IDF digunakan untuk mengetahui tingkat kemiripan kata dalam dokumen karya akhir mahasiswa. Pendekatan pemecahan masalah teknis ini diharapkan dapat memberikan solusi yang efektif dalam mengelola data administrasi tugas akhir dan mengurangi terjadinya plagiarisme serta kesalahan dalam penulisan tugas akhir mahasiswa dalam bahasa Indonesia di Politap.

2. METODE PENELITIAN

Penelitian ini bertujuan untuk pengembangan sistem deteksi plagiarisme tugas akhir mahasiswa berbasis Natural Language Processing menggunakan algoritma Jaro-Winkler dan TF-IDF. Metode yang digunakan dalam penelitian ini seperti metode pengumpulan data dan metode pengembangan perangkat lunak. Metode pengumpulan data terdiri dari studi literatur dan observasi. Sedangkan metode pengembangan sistem dalam penelitian ini mengadopsi Agile Software Development Method dengan menggunakan pendekatan model scrum[14]. Model scrum difokuskan pada tim yang terdiri dari beberapa anggota dengan peran yang berbeda-beda dalam mengembangkan perangkat lunak[15].

Pada studi literatur kegiatan yang dilakukan adalah mempelajari teori-teori pada buku dan jurnal ilmiah terkait pengembangan sistem deteksi plagiarisme, algoritma Jaro-Winkler, TF-IDF, serta teknologi Natural Language Processing (NLP). Observasi digunakan untuk melakukan penelusuran dan mengumpulkan judul-judul tugas akhir mahasiswa Politap yang telah disetujui oleh dosen pembimbing dan menjadikannya sebagai dataset pada penelitian ini.

2.1. Algoritma Jaro-Winkler Distance

Dalam algoritma Jaro-Winkler Distance, terdapat suatu rumus yang diterapkan untuk mengkomputasi jarak (d_j) antara dua string, yaitu S_1 dan S_2 .

$$d_j = \frac{1}{3} \times \left(\frac{m}{s_1} + \frac{m}{s_2} + \frac{m-t}{m} \right) \quad (1)$$

Dimana :

- m = Jumlah karakter yang sama persis
- $|S_1|$ = Panjang string 1
- $|S_2|$ = Panjang string 2
- t = Jumlah transposisi

Jarak teoritis antara dua karakter yang dianggap sama dapat dianggap benar jika tidak melebihi:

$$\left(\frac{\max(|S_1|, |S_2|)}{2} \right) - 1 \quad (2)$$

Akan tetapi bila mengacu kepada nilai yang akan dihasilkan oleh algoritma Jaro-Winkler Distance maka nilai maksimalnya adalah 1, yang menandakan kesamaan string yang di bandingkan mencapai seratus persen atau sama persis. Algoritma Jaro-Winkler Distance menggunakan prefix scale (p), menurut Winkler nilai standar untuk konstanta ini adalah $p = 0.1$. Dan prefix length (l) yaitu untuk menyatakan panjang prefix atau panjang karakter yang sama sampai ditemukan ketidaksetaraan (maksimum 4 karakter), maka untuk menghitung Jaro-Winkler Distance (d_w) adalah :

$$d_w = d_j + \left(l \times p(1 - d_j) \right) \quad (3)$$

Dimana :

- d_j = jaro distance untuk string S_1 dan String s_2
- l = panjang prefix (panjang karakter yang sama sebelum ditemukan ketidak setaraan) ilai maksimum 4 karakter
- p = konstanta scaling factor (nilai standar untuk konstanta ini menurut winkler adalah $p= 0.1$).

2.2. Term Frequency Inverse Document Frequency (TF-IDF)

Dalam perhitungan bobot menggunakan TF-IDF, dihitung terlebih dahulu nilai TF perkata dengan bobot masing-masing kata adalah 1. Sedangkan nilai IDF diformulasikan pada persamaan berikut.

$$Tf_{t,d} = \frac{\log tf+1}{0} \quad (4)$$

dimana setiap variabel dijelaskan sebagai berikut :

- tf = term frekuensi atau banyaknya kata pada dokumen
- $Tf_{t,d}$ = term frekuensi atau banyaknya kata t pada dokumen d atau pembobotan lokal.

Nilai IDF sebuah term (kata) dapat dihitung menggunakan persamaan berikut:

$$Idf_t = \log \left(\frac{N}{dft} \right) \quad (5)$$

- $Idft$ = inverse document-frequency atau pembobotan global
- N = banyaknya dokumen
- dft = banyaknya dokumen yang memiliki kata t .

Adapun algoritma yang digunakan untuk menghitung bobot ($W_{t,d}$) dengan mengalikan kedua persamaan sehingga menjadi:

$$w_{t,d} = Tf_{t,d} \times Idf_t \quad (6)$$

Dimana:

tft,d = term frekuensi atau banyaknya kata pada dokumen atau pembobotan lokal

$Idft$ = inverse document frequency atau pembobotan global

wt,d = nilai bobot akhir kata.

3. HASIL DAN PEMBAHASAN

3.1. Pembahasan

Data yang digunakan dalam penelitian ini merupakan data karya tugas akhir mahasiswa yang terdiri dari 20 dokumen Tugas Akhir Mahasiswa Jurusan Teknik Elektro dan Teknik Informatika Tahun 2022 yang diambil secara acak untuk digunakan sebagai Dataset seperti pada Tabel 1 berikut. Adapun data yang dikumpulkan adalah teks yang terdapat pada bagian judul dan deskripsi proposal tugas akhir.

Tabel 1. Perbandingan Algoritma A dan Algoritma B

No	Judul Tugas Akhir	Abstrak
1	Game Edukatif Untuk Pengenalan Permainan Tradisional Khas Daerah Indonesia Berbasis Android	Pengenalan Permainan Tradisional Khas Daerah Indonesia melalui Game Edukatif Berbasis Android merupakan sebuah langkah inovatif yang mengintegrasikan nilai-nilai tradisional Indonesia dengan teknologi kontemporer. Tujuan utama aplikasi ini adalah melestarikan permainan tradisional yang semakin terpinggirkan di era modern, dengan memanfaatkan multimedia. Aplikasi ini khusus dirancang untuk melibatkan generasi muda, khususnya anak-anak dan remaja, dalam proses pembelajaran dan pemahaman beragam permainan tradisional dari seluruh Indonesia. Dengan memberikan pengalaman interaktif yang edukatif dan menghibur, aplikasi ini tidak hanya memperkenalkan permainan tradisional, tetapi juga meningkatkan apresiasi terhadap keanekaragaman budaya serta pentingnya menjaga warisan budaya Indonesia bagi masa depan. Dalam pengembangan aplikasi ini, metode waterfall game edukatif digunakan, yang mencakup analisis kebutuhan, desain sistem, implementasi, dan pengujian. Berdasarkan hasil pengujian blackbox, aplikasi ini telah sukses melewati pengujian fungsionalitas, kinerja, kompatibilitas perangkat, penanganan kesalahan, serta antarmuka pengguna. Dengan demikian, penelitian ini menyimpulkan bahwa aplikasi game edukatif ini memberikan kontribusi dalam mempromosikan, melestarikan, dan memperkenalkan permainan tradisional khas daerah Indonesia kepada pengguna.
...
20	Aplikasi Pemesanan Tiket Speedboat Ketapang Dan Sukadana Berbasis Android	Speedboat adalah salah satu jenis transportasi laut yang dirancang khusus untuk beroperasi dengan kecepatan tinggi di perairan terbuka. Transportasi ini memiliki mesin yang kuat untuk mencapai kecepatan tinggi, sehingga cocok digunakan untuk penumpang, pariwisata, atau keperluan rekreasi. Seiring dengan perkembangan teknologi dan meningkatnya kebutuhan akan mobilitas, penggunaan speedboat semakin meluas dan mendapatkan banyak perhatian. Tujuan dari penelitian ini adalah membangun aplikasi pemesanan tiket speedboat Ketapang dan Sukadana berbasis android agar dapat memudahkan calon penumpang untuk membeli tiket tanpa harus datang jauh-jauh ke loket penjualan tiket. Metode penelitian yang digunakan dalam penelitian ini adalah Research and Development

No	Judul Tugas Akhir	Abstrak
		(R & D). Dalam penelitian terdiri dari beberapa rancangan, yaitu: rancangan arsitektur sistem, perancangan arus data, perancangan basis data, perancangan antarmuka, perancangan pengujian sistem dengan Black Box Testing. Adapun penelitian ini menghasilkan sebuah Aplikasi Pemesanan Tiket Speedboat Ketapang dan Sukadana Berbasis Android.

3.1.1. Algoritma Jaro Winkler

Berikut ini merupakan contoh penerapan perhitungan algoritma Jaro-Winkler Distance dari kata yang di ambil judul tugas akhir. Jika string S1 "Sistin Informasi Pengaduan" dan string S2 "Sistem Infromasi Penggaduan" maka : $m = 26.5$, $S1 = 26$, $S2 = 27$. Jumlah transposisi (jumlah karakter yang perlu ditukar posisinya untuk membuat keduanya cocok). Dalam hal ini, terdapat beberapa karakter yang perlu ditukar, yaitu "n" dan "m" pada kata "Informasi" dan "d" dan "g" pada kata "Pengaduan." Panjang string m (rata-rata panjang kedua string). Panjang "Sistim Informasi Pengaduan" adalah 26, dan panjang "Sistem Informasi Pengaduan" adalah 27. Jadi, $m = (26 + 27) / 2 = 26.5$.

Maka Nilai Jaro-D:

Jaro-D = (kesamaan karakter / panjang string 1 + kesamaan karakter / panjang string 2 + (kesamaan karakter - jumlah transposisi) / kesamaan karakter) / 3

Jaro-D = $(26 / 26 + 26 / 27 + (26 - 4) / 26) / 3$

Jaro-D = $(1 + 0.96296 + 0.84615) / 3$

Jaro-D = $2.80911 / 3 = 0.93637$

Jaro-Winkler Distance (Jaro-W):

Hitung Jaro-Winkler Scaling Factor (p). Misalnya, kita gunakan $p = 0.1$.

Jaro-W = $Jaro-D + (p * L * (1 - Jaro-D))$

Jaro-W = $0.93637 + (0.1 * 7 * (1 - 0.93637))$

Jaro-W = $0.93637 + (0.7 * 0.06363)$

Jaro-W = $0.93637 + 0.044541 = 0.98091$

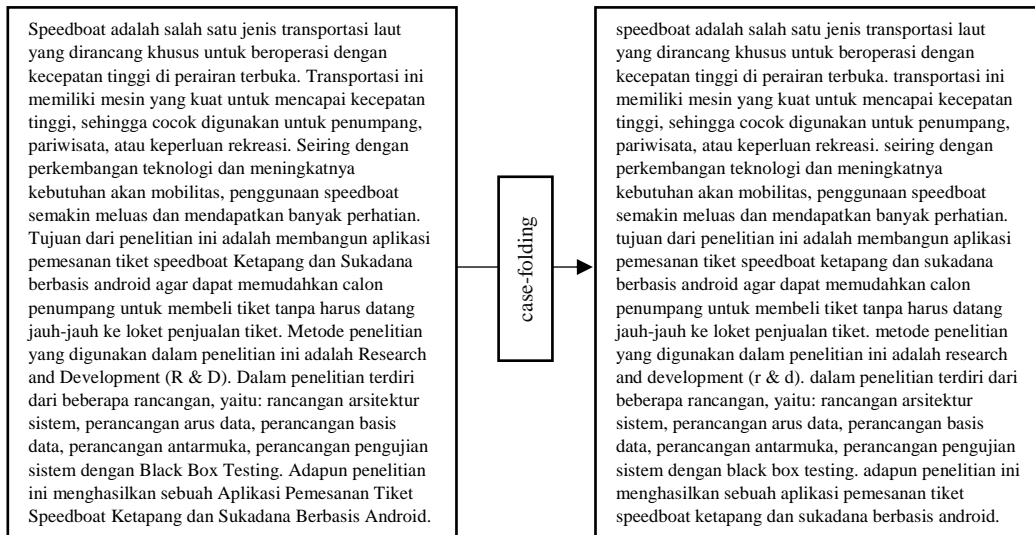
Hasil perhitungan Jaro-Winkler Distance antara "Sistin Informasi Pengaduan" dan "Sistem Informasi Pengaduan" adalah sekitar 0.98091. Nilai ini menunjukkan bahwa kedua string tersebut memiliki tingkat kemiripan yang sangat tinggi.

3.1.2. Algoritma TF-TDF

Langkah-langkah menggunakan algoritma TF-IDF (Term Frequency-Inverse Document Frequency) untuk menganalisis dokumen melibatkan pengolahan teks dan perhitungan TF-IDF seperti berikut.

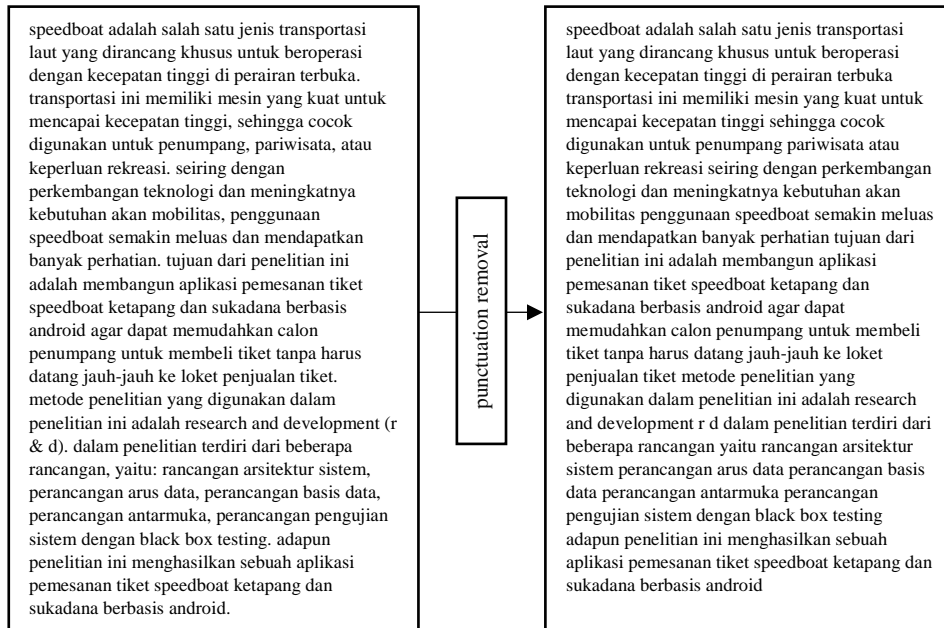
Pra-pemrosesan dilakukan dalam beberapa tahapan dengan tujuan untuk membersihkan kalimat dari karakter yang tidak diinginkan dan bisa mengganggu proses kalkulasi. Dalam penelitian ini dilakukan beberapa tahapan pra-pemrosesan yang terdiri dari:

1. Case-folding, merupakan sebuah teknik yang digunakan untuk mengubah seluruh karakter yang ada di dalam kalimat menjadi huruf kecil (lowercase). Transformasi kalimat yang telah melalui proses case-folding dapat dilihat pada Gambar 2.



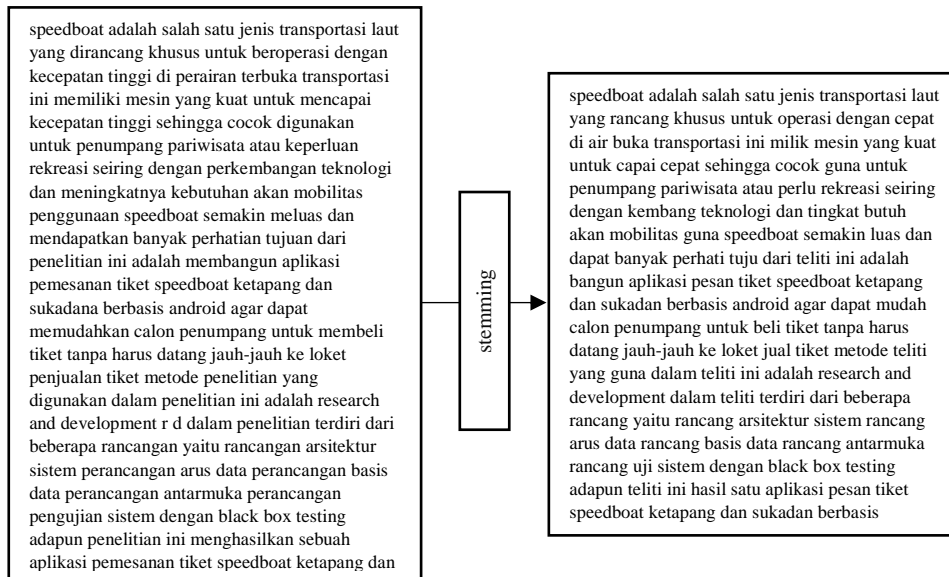
Gambar 2. Transformasi kalimat melalui proses case-folding

2. Punctuation removal, merupakan sebuah teknik yang akan menghapus seluruh karakter simbol, tanda baca dan karakter lain yang tidak memiliki arti di dalam kalimat. Transformasi kalimat yang telah melalui proses punctuation removal dapat dilihat pada Gambar 3.



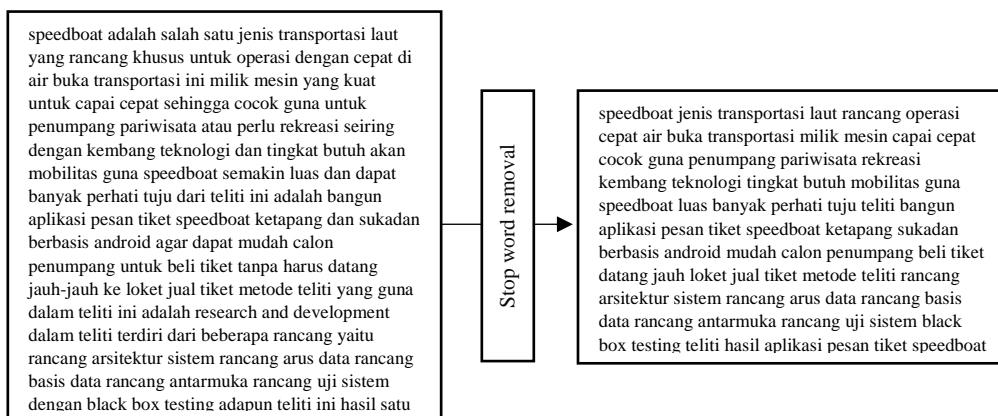
Gambar 3. Transformasi kalimat melalui proses punctuation removal

3. Proses stemming adalah langkah dalam pra-pemrosesan teks yang bertujuan untuk mengubah kata-kata ke bentuk dasarnya (kata dasar). Berikut adalah teks sebelum dan sesudah proses stemming dapat dilihat pada Gambar 4.



Gambar 4. Transformasi kalimat melalui proses stemming

4. Stop word removal, adalah teknik yang digunakan untuk menghapus kata-kata yang terdapat pada stop word list. Proses penghapusan stop words adalah langkah dalam pra-pemrosesan teks yang bertujuan untuk menghapus kata-kata umum atau kata-kata penghubung yang mungkin tidak memberikan informasi penting dalam analisis teks. Berikut adalah teks sebelum dan sesudah proses penghapusan stop words dapat dilihat pada Gambar 5.



Gambar 5. Transformasi kalimat melalui proses stopword removal

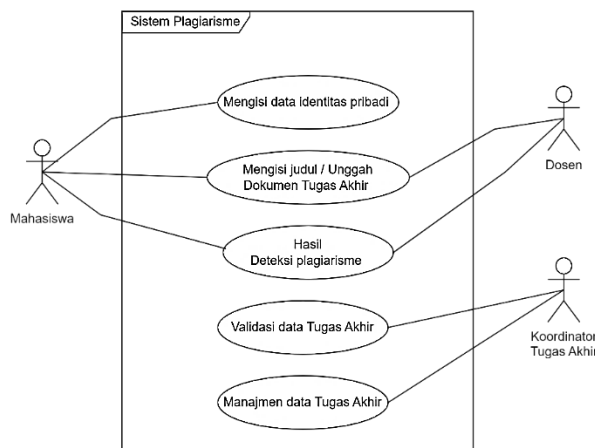
Setelah dilakukannya proses text mining dan menghasilkan kalimat yang memiliki peran penting didalam perhitungan, selanjutnya yaitu perhitungan TF- IDF. Proses menghitung nilai bobot, terlebih dahulu hitung jumlah kemunculan term pada setiap dokumen, kemudian menjumlahkan setiap kemunculan kata yang sama pada semua dokumen letakkan pada kolom DF. Selanjutnya, mencari IDF yaitu hitung jumlah dokumen dibagi DF kemudian hasil pembagian di-log- kan. Langkah selanjutnya, mencari WDT yaitu nilai TF dikalikan nilai IDF sehingga memperoleh nilai bobot term. Berikut hasil keseluruhan proses dari perhitungan TF-IDF yang disajikan pada Tabel 3.

Tabel 3. Hasil Pembobotan kata

Term	Tf	Df	N	Idf	Wdt
				$\log(n/df) + 1$	$Tf * Idf$
speedboat	4	1	20	2.301029996	9.204119983
transportasi	2	1	20	2.301029996	4.602059991
tinggi	2	5	20	1.602059991	3.204119983
penumpang	2	1	20	2.301029996	4.602059991
aplikasi	2	13	20	1.187086643	2.374173287

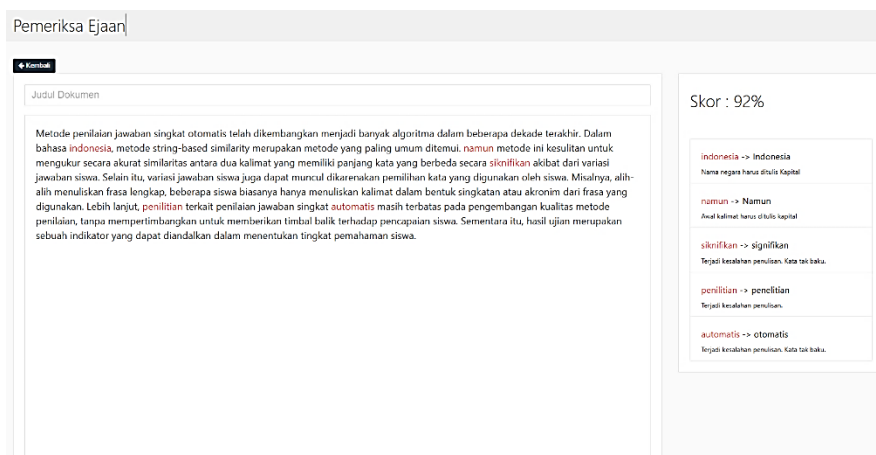
Term	Tf	Df	N	Idf	Wdt
				$\log(n/df) + 1$	$Tf * Idf$
...
tiket	4	1	20	2.301029996	9.204119983
android	2	6	20	1.522878745	3.045757491
research	7	15	20	1.124938737	7.874571156
development	10	15	20	1.124938737	11.24938737

Setelah melewati tahap perhitungan dengan algoritma Jaro-Winkler dan TF-IDF maka sistem yang dibangun untuk melakukan deteksi kesalahan penulisan dalam bahasa Indonesia dan deteksi plagiarisme tugas akhir mahasiswa digambarkan pada use case diagram seperti pada gambar 6 berikut.



Gambar 6. Use Case Diagram

Implementasi sistem deteksi kesalahan penulisan dalam bahasa Indonesia dan deteksi plagiarisme Tugas Akhir mahasiswa dapat dilihat pada gambar 7. Berikut. Dimana sistem akan secara otomatis melakukan deteksi pada kata yang salah atau tidak sesuai dengan penulisan bahasa Indonesia yang baik dan benar sesuai dengan panduan pada KBBI dan EYD. Kata yang salah akan diberikan penegasan dengan warna merah dan diberikan juga koreksi kata untuk perbaikannya.



Gambar 7. Form pengecekan kesalahan penulisan

Dengan begitu diharapkan mahasiswa dapat dengan mudah melakukan pengecekan penulisan karya akhir mereka sebelum dilakukan pembimbingan kepada dosen pembimbing, sehingga dapat meminimalisir kesalahan dan banyaknya koreksi yang kemungkinan akan diberikan. Sementara fitur yang digunakan untuk deteksi plagiarisme adalah seperti pada gambar 7 yang disajikan di bawah ini.

Tugas Akhir

PEMERIKSAAN KESALAHAN PENULISAN DAN KEMIRIPAN JUDUL TUGAS AKHIR

JUDUL YANG DIAZUKAN

Aplikasi Pemesanan Tiket Speedboat Ketapang Dan Sukadana Berbasis Android

DESKRIPSI / LATAR BELAKANG

Speedboat adalah salah satu jenis transportasi laut yang dirancang khusus untuk beroperasi dengan kecepatan tinggi di perairan terbuka. Transportasi ini memiliki mesin yang kuat untuk mencapai kecepatan tinggi, sehingga cocok digunakan untuk pemangung, pariwisata, atau keperluan militer. Selain dengan protokolnya yang ketat dan manajemennya yang efisien dan efektif, penggunaan speedboat memiliki risiko dan tantangan yang harus dihadapi. Tujuan dari penelitian ini adalah mengungkap aplikasi pemesanan tiket speedboat Ketapang dan Sukadana berbasis android agar dapat memudahkan calon penumpang untuk membeli tiket tanpa harus datang langsung ke loket penjualan tiket. Metode penelitian yang digunakan dalam penelitian ini adalah Research and Development (R & D). Dalam penelitian terdiri dari beberapa tahapan, yaitu: rancangan arsitektur sistem, perancangan basis data, perancangan antarmuka, perancangan pengujian sistem dengan Black Box Testing. Adapun penelitian ini menghasilkan sebuah Aplikasi Pemesanan Tiket Speedboat Ketapang dan Sukadana Berbasis Android.

Q Periksa

HASIL PEMERIKSAAN

NO	NAMA	AKSI	JUDUL	PERSENTASE KEMIRIPAN JUDUL	PERSENTASE KEMIRIPAN DESKRIPSI LATAR BELAKANG
1	DERI SARI ASTUTI		aplikasi pemesanan tiket speedboat ketapang dan sukadana berbasis android	100,00 %	34,59 %
2	DERI SARI ASTUTI		aplikasi pemesanan tiket speedboat bintang riki express 99 ketapang berbasis android	47,38 %	15,57 %
3	ANGGI JUNIITA		SISTEM INFORMASI LOKET PENJUALAN TIKET KAPAL PADA PT. DHARMA LAUTAN UTAMA BERBASIS WEB DI KETAPANG	16,67 %	20,58 %
4	DWI RAMADHANI		Aplikasi Pemesanan Air Mineral di Kabupaten Ketapang berbasis android	19,83 %	11,89 %
5	RICO PAHLERI		APLIKASI PEMESANAN AMBULANS DI KABUPATEN KETAPANG BERBASIS ANDROID	23,11 %	7,09 %

Gambar 7. Form Pengecekan Plagiarism Tugas Akhir

Pada gambar 7. Form pengecekan plagiarisme tugas akhir dapat digunakan oleh mahasiswa sebagai sarana pengecekan tugas akhir untuk mendapatkan gambaran seberapa tingkat kemiripan dengan tugas akhir yang sudah pernah dilakukan oleh penelitian terdahulu. Tingkat kemiripan disajikan dalam bentuk presentase agar lebih mudah untuk dipahami oleh mahasiswa maupun pengguna lainnya. Dengan fitur tersebut diharapkan tingkat plagiarism tugas akhir mahasiswa pada Politeknik Negeri Ketapang dapat diatasi dengan baik.

4. KESIMPULAN

Sistem yang telah dibangun mengalami serangkaian pengujian, termasuk uji black-box testing dan Uji Coba Aplikasi (UAT). Selain itu, sistem deteksi plagiarisme juga menjalani sejumlah pengujian hingga diperoleh hasil pengujian yang paling akurat. Sistem ini mampu beroperasi dengan baik dan memberikan umpan balik yang tepat, baik dalam mengoreksi kesalahan penulisan maupun dalam mengecek tingkat kemiripan tugas akhir mahasiswa. Sebagai rekomendasi bagi peneliti berikutnya, sebaiknya dilakukan penelitian untuk menguji efektivitas algoritma Jaro-Winkler dan TF-IDF dengan berbagai algoritma lainnya. Setelah menemukan algoritma yang paling efektif, pengembangan sistem dapat dilanjutkan dengan mengadopsi algoritma tersebut.

DAFTAR PUSTAKA

- [1] Jambi KM, Khan IH, Siddiqui MA. Evaluation of Different Plagiarism Detection Methods: A Fuzzy MCDM Perspective. *Applied Sciences*. 2022 Apr 30;12(9):4580.
- [2] Swari MHP, Putra CA, Handika IPS. Plagiarsm Checker pada Sistem Manajemen Data Tugas Akhir. *JSI*. 2021 Dec 5;7(2):192–201.
- [3] Dillak RY, Laumal F, Kadja LJ. Sistem Deteksi Dini Plagiarisme Tugas Akhir Mahasiswa Menggunakan Algoritma Ngrams Dan Winnowing. *Flash*. 2016 Jun 13;2(1):12.
- [4] Subadri A, Pratama I. Sistem Deteksi Plagiarism Pada Judul Tugas Akhir Menggunakan Metode Rabin-Karp Berbasis Web. *Technologia*. 2022 Oct 15;13(4):306.
- [5] Aritomatika EJ, Sanjaya A, Widodo DW. Implementasi Algoritma Jaro Winkler Distance Untuk Pendeteksi Kesamaan Kata Dalam Pengembangan Aplikasi English Conversation. 2021.
- [6] Tannga MJ, Rahman S. Analisis Perbandingan Algoritma Levenshtein Distance Dan Jaro Winkler Untuk Aplikasi Deteksi Plagiarisme Dokumen Teks.

- [7] Prasetyo A, Baihaqi WM, Had IS. Algoritma Jaro-Winkler Distance: Fitur Autocorrect dan Spelling Suggestion pada Penulisan Naskah Bahasa Indonesia di BMS TV. *JTIK*. 2018 Oct 1;5(4):435.
- [8] Nurjannah M, Astuti IF. Penerapan Algoritma Term Frequency-Inverse Document Frequency (TF-IDF) Untuk Text Mining.
- [9] Amrizal V. Penerapan Metode Term Frequency Inverse Document Frequency (TF-IDF) Dan Cosine Similarity Pada Sistem Temu Kembali Informasi Untuk Mengetahui Syarah Hadits Berbasis Web (Studi Kasus: Hadits Shahih Bukhari-Muslim). *J Teknik inform*. 2018 Nov 28;11(2):149-64.
- [10] P. Novantara, "Implementasi Algoritma Jaro-Winkler Distance Untuk Sistem Pendeteksi Plagiarisme Pada Dokumen Skripsi," *BI*, vol. 3, no. 1, Apr 2018, doi: [10.25134/buffer.v3i2.960](https://doi.org/10.25134/buffer.v3i2.960).
- [11] R. Y. Dillak, F. Laumal, dan L. J. Kadja, "Sistem Deteksi Dini Plagiarisme Tugas Akhir Mahasiswa Menggunakan Algoritma Ngrams Dan Winnowing," *Flash*, vol. 2, no. 1, hlm. 12, Jun 2016, doi: [10.32511/jiflash.v2i1.19](https://doi.org/10.32511/jiflash.v2i1.19).
- [12] T. M, H. Haripuddin, dan S. Sanatang, "Pengembangan Sistem Deteksi Kemiripan TA Berbasis Algoritma Rabin Karp & Levensthein Naskah Skripsi Mahasiswa JTIK," *ETNIK*, vol. 2, no. 2, hlm. 131-140, Feb 2023, doi: [10.54543/etnik.v2i2.155](https://doi.org/10.54543/etnik.v2i2.155).
- [13] Z. A. Rahmadiani, D. Priharsari, dan A. R. Perdanakusuma, "Analisis Persepsi Dosen Terhadap Kebijakan Penggunaan Turnitin untuk Mendeteksi Plagiarisme Skripsi Mahasiswa", *J-PTIHK*, vol. 7, no. 3, hlm. 1265-1272, Mar 2023.
- [14] Pressman, R.S. (2010), *Software Engineering : a practitioner's approach*, McGraw - Hill, New York
- [15] Sommerville, Ian. 2011. *Software Engineering (Rekayasa Perangkat Lunak)*. Jakarta : Erlangga.