

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Dalam konteks kehidupan akademis, integritas ilmiah merupakan fondasi utama yang menjamin kualitas dan keaslian dalam penelitian. Salah satu ancaman serius terhadap integritas ini adalah Tindakan *plagiarisme* yang semakin menjadi perhatian dalam lingkungan akademis ini. *Plagiarisme* dapat merugikan kepercayaan Masyarakat terhadap hasil penelitian, mencoreng nama baik penelitian, dan menghancurkan kejujuran dan keaslian penelitian dalam menghasilkan pengetahuan baru.[1].

Tingginya tekanan dalam dunia akademis, Bersama dengan kemudahan akses informasi melalui internet telah memberikan peluang yang begitu besar untuk terjadinya *plagiarisme*. Peneliti, mahasiswa, dan akademis sering kali terlibat dalam kegiatan menyalin ide atau teks tanpa memberikan kredit yang seharusnya tidak menghancurkan nilai keaslian dan integrasi karya ilmiah.[2].

Dengan semakin kompleksnya bentuk-bentuk *plagiarisme* seperti *paraphrasing* atau pengganti kata, metode-metode baru yang digunakan oleh para pelaku *plagiarisme*, diperlukan suatu sistem deteksi *plagiarisme* yang mampu mengidentifikasi dan menganalisis Tingkat *plagiarisme* dengan akurat. Oleh karena itu, penelitian ini bertujuan untuk mengembangkan suatu sistem deteksi plagiarisme pada kajian ilmiah dengan tujuan utama untuk meningkatkan integritas akademis. Sistem ini diharapkan dapat memberikan perlindungan efektif terhadap resiko plagiarisme, mendeteksi Tindakan plagiat dalam berbagai

bentuk dan memberikan dasar yang kuat bagi Lembaga Pendidikan untuk menjaga kualitas dan integritas penelitian yang dihasilkan.

Perpustakaan sebagai pusat penyimpanan dan distribusi informasi ilmiah yang memiliki tanggung jawab besar untuk mendukung Upaya mencegah plagiarisme. Namun, perpustakaan yang bergantung pada pengolahan manual dokumen fisik dan digital sering kali menghadapi kesulitan dalam mendeteksi *plagiarisme* secara efektif. Adapun beberapa tantangan yang di hadapi seperti jumlah dokumen yang banyak, format beragam, kurangnya sistem terintegrasi, perkembangan teknik *plagiarisme*, dan minimnya kesadaran pengguna. Berdasarkan permasalahan tersebut diperlukan suatu sistem deteksi plagiarisme yang otomatis dan terintegrasi dengan baik dalam pengolahan data di perpustakaan.

Sistem deteksi *plagiarisme* yang ada saat ini, seperti *Turnitin* dan *Grammarly*, sudah banyak digunakan secara luas. Namun, beberapa di antaranya memerlukan biaya berlangganan yang tidak terjangkau bagi sebagian besar individu atau institusi kecil. Selain itu, tidak semua sistem yang ada mendukung bahasa Indonesia secara optimal, yang menyebabkan kurangnya akurasi dalam mendeteksi *plagiarisme* pada karya ilmiah yang ditulis dalam bahasa tersebut.

Metode *Cosine Similarity* adalah salah satu pendekatan yang banyak digunakan dalam mendeteksi kesamaan teks. Metode ini bekerja dengan menghitung tingkat kemiripan antara dua dokumen. Perhitungan metode ini didasarkan pada dua vektor yang mewakili kata-kata dalam dokumen yang dibandingkan. Kelebihan dari metode ini adalah kemampuan dalam mengukur

kemiripan antar dokumen dengan efisiensi komputasi yang tinggi, sehingga cocok diimplementasikan dalam sistem berbasis *website*. [3].

Namun, tantangan utama dalam *implementasi* metode ini terletak pada penanganan variasi Bahasa, seperti penggunaan *sinonim* atau *parafase* yang sering kali tidak dapat diidentifikasi oleh *metode cosine similarity* yang hanya mengukur kesamaan berdasarkan *frekuensi* kata. Oleh karena itu diperlukan pendekatan tambahan seperti penggunaan model *TF-IDF* untuk meningkatkan keakuratan sistem dalam mendeteksi *plagiarisme* pada karya ilmiah.

Melihat pentingnya deteksi plagiarisme dalam menjaga kualitas karya ilmiah, penelitian ini bertujuan untuk mengembangkan sebuah sistem deteksi *plagiarisme* berbasis *website* menggunakan metode *cosine similarity*. Sistem ini diharapkan dapat menjadi alternatif yang terjangkau, mudah diakses, dan efektif untuk mendeteksi kesamaan teks, terutama pada karya ilmiah yang ditulis menggunakan bahasa Indonesia.

Melalui pengembangan sistem deteksi *plagiarisme* pada kajian ilmiah berbasis *website* ini, diharapkan menjadi solusi yang lebih mudah diakses, dan efektif untuk mendeteksi *plagiarisme* pada karya ilmiah, khususnya yang ditulis dengan menggunakan bahasa Indonesia. Sistem ini diharapkan dapat membantu institusi Pendidikan, penelitian, dan mahasiswa dalam menjaga keaslian karya ilmiah mereka dan menjaga terjadinya pelanggaran etika akademik.

## 1.2 Tujuan dan Manfaat

### 1.2.1. Tujuan

Berikut tujuan dari penelitian dibuatnya sistem deteksi *plagiarisme* pada kajian ilmiah berbasis *website*:

1. Sistem ini dirancang untuk mendeteksi suatu karya ilmiah atau teks memiliki kemiripan yang signifikan pada karya lain yang sudah ada.
2. Sistem ini juga berfungsi sebagai alat pencegahan terhadap tindakan penyalinan atau penjiplakan karya ilmiah orang lain tanpa memberikan distribusi yang sesuai.
3. Kualitas penelitian dapat ditingkatkan karena penulis akan lebih berhati-hati dalam memastikan bahwa karya mereka adalah hasil orisinal dan bukan hasil penyalinan dari sumber lain.
4. Sistem berbasis website ini memudahkan pengguna seperti dosen, peneliti, mahasiswa atau editor jurnal, untuk melakukan pengecekan plagiarisme secara online dan efisien tanpa perlu mengunduh software khusus atau pengecekan manual.
5. Metode cosine similarity ini digunakan untuk mengukur dua dokumen atau teks dengan memberikan nilai yang menunjukkan tingkat kesamaan. Nilai ini bisa membantu peneliti atau pengajar dalam menentukan apakah dua karya memiliki kesamaan yang mencurigakan.

### 1.2.2. Manfaat

Berikut manfaat yang didapatkan dari penelitian ini pada politeknik harapan bersama:

1. Meningkatkan kualitas kajian ilmiah, sistem deteksi *plagiarisme* yang handal dapat membantu meningkatkan kualitas kajian ilmiah dengan memastikan bahwa setiap karya memiliki nilai yang tinggi.
2. sistem deteksi ini *plagiarisme* ini dapat menjadi alat yang mendukung proses penilaian kajian ilmiah. Dengan memberikan informasi detail tentang potensi *plagiarisme*.

Manfaat pada mahasiswa berikut ini:

1. Meningkatkan kesadaran akan plagiarisme bahwa sadar akan bahaya dan konsekuensi plagiarisme.
2. Meningkatkan keterampilan dalam menulis seperti parafrase, Menyusun kutipan, dan mengintegrasikan sumber secara etis.
3. Motivasi untuk kreativitas dan inovasi.
4. Penyediaan alat bantu dalam menyusun karya untuk memeriksa keaslian karya mereka sebelum diserahkan.

Manfaat pada masyarakat:

1. Meningkatkan karya asli sistem ini membantu melindungi hak cipta dan hak kekayaan intelektual sehingga pencipta asli dapat menerima dan pengakuan yang layak atas karyanya.
2. Meningkatkan kepercayaan pada hasil penelitian

3. Sistem ini mendorong Masyarakat untuk menghargai integritas intelektual dan etika dalam menciptakan karya, baik dalam lingkungan akademis maupun professional.

### 1.3 Tinjauan Pustaka

Pada penyusunan laporan ini terdapat penelitian sebelumnya yang berkaitan dengan latar belakang masalah yang menjadi referensi dalam pembuatan laporan yang dibuat. Berikut adalah beberapa penelitian terkait mengenai aplikasi sistem deteksi *plagiarisme* pada kajian ilmiah antara lain berikut:

Artikel pertama berjudul "Perbandingan Hasil Deteksi Plagiarisme Dokumen dengan Metode *Jaro-Winkler Distance* dan Metode *Latent Semantic Analysis*" membahas perbandingan dua metode deteksi *plagiarisme*. Metode *Jaro-Winkler Distance* mengukur kesamaan antar *string* berdasarkan urutan kata, sementara *Latent Semantic Analysis* fokus pada kata-kata dalam dokumen tanpa memperhatikan karakter linguistik. Hasil penelitian menunjukkan bahwa metode *Jaro-Winkler Distance* memberikan hasil deteksi *plagiarisme* yang lebih akurat, mencapai 100% pada data yang sama persis, dibandingkan dengan 97,14% pada metode *Latent Semantic Analysis*. [4].

Artikel berjudul "Deteksi Plagiarisme Tingkat Kemiripan Judul Skripsi pada Fakultas Teknologi Informasi Menggunakan *Algoritma Winnowing*" membahas penggunaan *algoritma Winnowing* untuk menentukan kemiripan teks judul skripsi. *Algoritma* ini mendeteksi kesamaan kata atau kalimat (*common subsequence*) dalam dua atau lebih teks dengan menggunakan

konsep *fingerprint*, yang diperoleh melalui proses *hashing n-gram*. Sistem ini memungkinkan Ketua Program Studi atau Koordinator Tugas Akhir untuk memasukkan judul skripsi yang diajukan, lalu sistem secara otomatis mengecek dan menampilkan hasil kemiripan, yang dapat digunakan sebagai dasar pengambilan keputusan apakah judul skripsi tersebut diterima atau ditolak.[5].

Artikel yang ketiga, berjudul "Sistem Deteksi *Plagiarisme* Lintas Bahasa Menggunakan Algoritma *TF-IDF*," menjelaskan bahwa plagiarisme adalah tindakan mengambil ide atau teks orang lain dan mengakuinya sebagai milik sendiri. Penelitian ini bertujuan untuk mengurangi plagiarisme dengan menggunakan *algoritma TF-IDF*. Prosesnya meliputi *translasi, tokenizing, eliminasi, stemming*, dan perhitungan *cosine similarity* untuk mendapatkan persentase kesamaan dokumen. Hasil pengujian menunjukkan bahwa sistem ini efektif dalam mendeteksi kesamaan dokumen yang diuji.[6].

Artikel keempat, berjudul "Sistem Deteksi *Plagiarisme* pada Judul Tugas Akhir Menggunakan Metode *Rabin-Karp* Berbasis Web," menjelaskan bahwa perkembangan teknologi informasi mempermudah pertukaran informasi, namun juga meningkatkan risiko *plagiarisme*. Artikel ini membahas perlunya sistem untuk memeriksa *plagiarisme* pada judul tugas akhir mahasiswa secara cepat dan tepat menggunakan algoritma *Rabin-Karp*, yang unggul dalam pencarian string dengan pola panjang. Prosesnya melibatkan *text preprocessing (case folding, filtering, tokenizing)* sebelum menggunakan algoritma *Rabin-Karp* untuk menghitung persentase kemiripan judul. Pengujian menunjukkan

akurasi similaritas sebesar 70%. Algoritma ini dianggap cukup layak sebagai metode sederhana untuk mendeteksi *plagiarisme*, dengan tujuan utama mendeteksi dan mencegah praktik plagiarisme sedini.[7].

Artikel kelima, berjudul "Analisis Kinerja *Algoritma Levenshtein Distance* dalam Mendeteksi Kemiripan Dokumen Teks," membahas peningkatan *plagiarisme* pada dokumen teks. Plagiarisme umumnya dilakukan dengan mengubah struktur kata dan kalimat, serta menyisipkan, menghapus, atau mengganti kata. Penelitian ini membangun sistem yang menggunakan algoritma *Levenshtein distance* dengan *case folding*, *tokenizing*, *stopword removal*, *stemming*, dan *sorting* untuk mendeteksi kemiripan antar dokumen teks. Algoritma ini menghasilkan nilai *distance* yang menentukan persentase kesamaan. Analisis menunjukkan bahwa *sorting* sangat berpengaruh pada kinerja algoritma. Hasil terbaik untuk data set 1 (pengubahan struktur kata) diperoleh dengan *stopword removal*, *stemming*, dan *sorting*. Hasil terbaik untuk data set 2 (pengubahan struktur kalimat) diperoleh dengan *stopword removal*, *stemming*, dan *sorting*. Hasil terbaik untuk data *real* (tanpa pengubahan) diperoleh dengan *stemming* dan *sorting*[8].

**Tabel 1. 1** GAP Penelitian

No	Judul	Metode	Perbedaan
1	Perbandingan Hasil Deteksi Plagiarisme Dokumen dengan Metode Jaro-Winkler Distance dan Metode Latent Semantic Analysis	Metode jaro-winker distance dan Metode Latent Semantic Analysis	Bahwa metode Jaro-Winkler Distance memberikan hasil deteksi plagiarisme yang lebih akurat, mencapai 100% pada data yang sama persis, dibandingkan dengan 97,14% pada metode Latent Semantic Analysis.
2	Deteksi Plagiarisme Tingkat Kemiripan Judul Skripsi pada Fakultas Teknologi Informasi Menggunakan Algoritma WInnowing.	Menggunakan Algoritma WInnowing	Menggunakan konsep fingerprint yang diperoleh dengan hasil n-gram.

3	<p>Sistem Deteksi Plagiarisme Lintas Bahasa</p> <p>Menggunakan Algoritma <i>TF-IDF</i>,</p>	<p>Menggunakan algoritma <i>TF-IDF</i></p>	<p>Prosesnya meliputi translasi, tokenizing, eliminasi, stemming, dan perhitungan <i>cosine similarity</i> untuk mendapatkan persentase kesamaan dokumen.</p>
4	<p>Sistem Deteksi Plagiarisme pada Judul Tugas Akhir</p> <p>Menggunakan Metode <i>Rabin-Karp</i> Berbasis Web</p>	<p>Menggunakan metode <i>Rabin-Karp</i></p>	<p>Berbasis <i>website</i></p>
5	<p>Analisis Kinerja Algoritma <i>Levenshtein Distance</i> dalam Mendeteksi Kemiripan Dokumen Teks</p>	<p>Menggunakan algoritma <i>Levenshtein Distance</i></p>	<p>Hasil terbaik untuk data real (tanpa pengubahan) diperoleh dengan stemming dan sorting.</p>

#### **1.4 Data Penelitian**

Dalam penelitian ini data yang diambil dari Google Scholar yang merupakan salah satu repositori besar untuk publikasi ilmiah. Dalam Google Scholar menyediakan berbagai jenis dokumen akademik seperti artikel jurnal, tesis. Pemilihan google scholar sebagai sumber data dengan cakupan luas dari publikasi ilmiah yang diindeks. Data yang diambil berdasarkan topik tertentu seperti judul, abstrak dan berbentuk file dengan format PDF. Dengan penelitian ini dapat mengakses informasi terbaru dan terkemuka di bidangnya yang membantu dalam memastikan validasitas dan keandalan hasil penelitian.