

BAB I

PENDAHULUAN

1.1 Latar Belakang

Teknologi berkembang begitu pesat sehingga berdampak pada setiap aspek kehidupan manusia, termasuk pendidikan. Manusia kini bisa lebih mudah mengakses internet untuk meneliti berbagai topik ilmiah berkat teknologi. Plagiarisme adalah salah satu contoh umum penyalahgunaannya. Plagiarisme berasal dari dua kata Latin, yang berarti *plagiarius* penculik, dan *plagiare* yang berarti mencuri. Penggunaan kata-kata atau ide orang lain tanpa memberikan penghargaan kepada penulis aslinya seolah-olah karya atau ide tersebut adalah miliknya sendiri disebut plagiarisme. Selain dilema moral, plagiarisme akan membuat mahasiswa menjadi kurang kreatif dalam memasukkan pemikiran dan pendapatnya ke dalam informasi yang ingin disajikan. Hal ini kemungkinan akan terus dilakukan karena belum adanya sistem yang digunakan untuk mengidentifikasi tingkat kemiripan teks pada instansi terkait.

Dengan menilai jurnal yang dapat mengidentifikasi kesamaan beberapa dokumen dengan membandingkan isinya, maka permasalahan plagiarisme dapat diatasi. Proses ini menghasilkan nilai atau bobot kemiripan jurnal yang telah dibandingkan. Sebab itu, sangat penting untuk memiliki mekanisme yang efektif dan handal dalam mendeteksi kemiripan teks pada dokumen jurnal yang diunggah, guna mencegah dan mengurangi tindakan plagiarisme yang merugikan bagi proses pembelajaran dan penilaian. Penggunaan teknologi dan metode deteksi plagiarisme yang canggih menjadi sangat relevan dan diperlukan. Salah satu metode yang telah terbukti efektif dalam mendeteksi kemiripan teks adalah metode *cosine similarity*

(Nugroho et al., 2021). Untuk mencegah *Cosine similarity* dalam memahami makna kalimat, vektor ini didasarkan pada jumlah kata yang serupa antara dua dokumen teks. Karena pendekatan *Cosine similarity* mempunyai nilai akurasi yang lebih besar dibandingkan dengan metode *Jaccard similarity*, maka peneliti menggunakannya. Hal ini disebabkan karena pendekatan *Cosine similarity* membandingkan N-gram yang sejajar satu sama lain dari dua perbandingan, sehingga menyamakan panjang vektor data Nurdiana et al. dalam (Rahmata, 2020). Manfaat utama sistem ini adalah kemampuannya untuk membandingkan jurnal yang diunggah dengan jurnal yang sudah disimpan dalam sistem untuk menentukan seberapa mirip jurnal tersebut.

Berdasarkan uraian di atas, Penelitian ini bertujuan untuk mengembangkan sistem yang dapat mengidentifikasi kemiripan teks secara otomatis dengan memanfaatkan algoritma *cosine similarity*. Proses pengembangan melibatkan pengumpulan data jurnal, pemrosesan teks, perhitungan *cosine similarity*, dan penyajian hasil kepada pengguna. Sistem ini dirancang untuk memberikan informasi akurat mengenai tingkat kemiripan jurnal yang dibandingkan.

1.2 Rumusan Masalah

Bagaimana membangun sistem informasi Perancangan sistem informasi identifikasi kesamaan teks berbasis web dengan menggunakan *algoritma cosine similarity* ?

1.3 Batasan Masalah

1. Dokumen yang digunakan hanya berekstensi (extension) pdf pada sistem identifikasi kemiripan teks berbasis web.
2. Sistem tidak dapat mendeteksi persamaan kata yang ada pada dokumen.

3. Pengecekan plagiarisme jurnal dengan jumlah data 200 artikel.

1.4 Tujuan Penelitian

Tujuan pada penelitian ini adalah merancang sistem yang digunakan untuk perancangan sistem informasi identifikasi kemiripan teks berbasis web menggunakan algoritma *cosine similarity*.

1.5 Manfaat penelitian

1. Mengetahui bahwa metode *cosine similarity* dapat digunakan untuk mengidentifikasi kemiripan teks pada dokumen dalam bentuk teks.
2. Diharapkan dapat memudahkan dalam mengidentifikasi kemiripan teks secara terkomputerisasi.
3. Dapat dijadikan referensi untuk pengembangan sistem yang berkelanjutan.