

BAB I

PENDAHULUAN

I.1 Latar Belakang

Plagiarisme atau penyalahgunaan kekayaan intelektual, serta penyalinan teks secara substansial atas karya lain. Produk intelektual yang berupa ide, data, dan tulisan merupakan bentuk karya yang sering dijiplak. Plagiarisme dapat terjadi baik disengaja maupun tidak disengaja, kedua alasan yang menjadi pendorong tersebut tetap dianggap plagiarisme apabila terdapat kesamaan pada dua makalah penelitian tanpa mengutip teks asli dan menggantinya dengan kata-kata sendiri (Shadiqi, 2019). Plagiarisme hanyalah menyalin dan menempelkan karya orang lain tanpa memberikan penghargaan kepada penulisnya, penemu, atau penggagas asli. Karya ilmiah merupakan suatu tulisan ataupun karangan yang didapatkan sesuai dengan sifat keilmuannya dan didasari dari berbagai hasil pengamatan, penelitian, dan peninjauan terhadap bidang ilmu tertentu, pembuatan karya ilmiah tidak sedikit orang yang membuat dan butuh waktu yang sangat banyak, Kebanyakan orang menjiplak (copy paste) karya orang lain (Marsha, 2021). Hal ini disebut sebagai tindakan plagiarism.

Plagiarism berasal dari bahasa inggris *plagiary* dan bahasa latin *plagiarius* yang memiliki arti penjiplak atau pencuri. Sehingga plagiarisme memiliki makna yaitu perbuatan menjiplak ide atau karya seseorang yang akan diakui sebagai karya

sendiri tanpa menyebutkan sumber ide atau gagasan tersebut berasal(Pahlevi R, 2022). Plagiarisme atau sering disebut dengan plagiat adalah tindakan dimana seorang penulis mengambil sebuah teori, pendapat, opini, dan lain-lain dari hasil pemikiran orang lain dan menjadikannya seolah seperti teori mereka sendiri(Dwi Yulianto, L., Gata, W., Frieyadie, & Saputra, 2022). Perkembangan teknologi informasi pada saat ini berkembang begitu pesat, dengan adanya teknologi dapat membantu atau meringankan pekerjaan manusia. Selain itu berkembangnya suatu teknologi tentu ada dampak positif dan negatifnya, salah satu contoh teknologi yang membantu aktivitas manusia adalah sistem informasi dan negatifnya adalah orang-orang cenderung sesuatu yang instan, bentuknya yaitu seperti plagiat (Mubarak, 2022).

Winnowing merupakan algoritma yang dipergunakan pada proses *document fingerprinting* dengan memanfaatkan fungsi *hashing* dari teks. Pada algoritma *winnowing* dilakukan prosesnya perhitungan nilai *hash* setiap *k-gram*, guna mendapatkan nilai *hash* kemudian menggunakan fungsi *rolling hash*. Setelah itu di bentuk *window* dari nilai *hash* tersebut. Pada setiap *window* kemudian dipilih nilai *hash* yang terkecil. Jika terdapat lebih dari satu nilai *hash* yang terkecil, maka nilai *hash* yang paling kanan yang digunakan untuk membentuk *window*. Setelah itu seluruh nilai *hash* yang dipilih kemudian disimpan dijadikan *fingerprint* dari dokumen. *Fingerprint* ini nantinya yang akan dijadikan dasar pembandingan kesamaan antara teks yang telah dimasukan(Yuda, 2022).

Penelitian oleh(Dhamayanti & Sari, 2019) Bertujuan memberikan solusi pada permasalahan yang dihadapi salah satu instansi dengan cara membangun aplikasi pendeteksi plagiasi menggunakan metode *cosine similarity*. *Cosine similarity* digunakan untuk menghitung nilai kemiripan dengan menyamakan kata perkata dan menjadi salah satu teknik untuk mengukur kemiripan teks dengan tingkat akurasi yang tinggi. Selanjutnya dilakukan oleh(Budiman & Widjaja, 2020) menganalisis kemiripan tertinggi dengan algoritma *string similarity cosine similarity* kombinasi preprocessing, hasil percobaan menunjukkan penggunaan menghasilkan nilai distribusi yang paling tinggi dan rasio kemiripan. Penelitian yang lain(Wahyuni et al., 2017) menggunakan metode *cosine similarity* dalam normalisasi panjang vektor data dengan membandingkan N-gram yang sejajar satu sama lain dari 2 pembanding yang bertujuan membuat sebuah sistem berbasis web yang dapat mengklasifikasikan dokumen secara otomatis menggunakan algoritma cosine similarity dalam proses pengklasteran menggunakan pembobotan TF-IDF, namun pada hasil pembahasan tidak memberikan hasil klasifikasi pembobotan TF-IDF.

Jaccard Kemiripan merupakan algoritma yang membandingkan dua dokumen dengan menghitung persamaan atau perbedaan antara dokumen-dokumen tersebut. Untuk melakukan kemiripan N-gram dan jaccard diperlukan suatu algoritma yang berperan sebagai sidik jari dokumen dan digunakan untuk mendukung algoritma penampi. (Sunardi et al., 2018). Penelitian lain(Desena & Solichin, 2021) menggunakan metode algoritma menampi N-gram dengan metode kesamaan Jaccard. N-gram berpengaruh besar terhadap hasil skor kemiripan yang memerlukan n jumlah blok huruf. Jaccard Kemiripan merupakan algoritma yang

membandingkan dua dokumen dengan menghitung kemiripan teks/dokumen. Terkait dari hasil penelitian yang dilakukan penggunaan nilai N-gram perlu di analisis kembali untuk keakuratan similarity dari metode yang digunakan.

Beberapa penelitian sebelumnya yang menjadi rujukan penelitian ini, diantaranya Penelitian yang judul penelitian "*Implementation of the Winnowing Algorithm in Detecting Plagiarism in Title and Abstract of Student's Final Project*". Penelitian ini menggunakan algoritma winnowing deteksi kesamaan abstrak dan judul penelitian pada tugas akhir, dengan dilakukan pengecekan terhadap sampel judul dan abstrak yang telah diunduh dari Google, kemudian Hasil perhitungan tingkat kesamaan dapat diakui sebagai tolok ukur penerimaan judul atau abstrak penelitian. Kesimpulan yang dihasilkan adalah penerapan metode winnowing dapat membantu dalam menemukan hasil yang serupa dari judul dan abstrak tugas akhir siswa dengan skenario *run test*, yaitu $n = 7$ dengan persentase 3,07% berkaitan setelah di lakukan pengujian(Nasien D, 2022). Penelitian selanjutnya dengan judul "*Tetun Language Plagiarism Detection With Text Mining Approach Using N-gram and Jaccard Similarity Coefficient*" Penelitian ini mengekstrak dokumen teks, selanjutnya menggunakan metode n-gram untuk mengambil huruf dari beberapa string dari sebuah kata untuk mengakhiri dokumen dengan 4-gram. Hasil penelitian ini adalah dapat menunjukkan kinerja sistem pendeteksi plagiarisme untuk semua kategori dengan persentase kesamaan di atas 98,99%(Costa & Mali, 2021).

Penelitian selanjutnya dengan judul "Deteksi Plagiarisme Berbasis Parafrase Pada Teks Bahasa Indonesia". Studi ini menyelidiki algoritma pembelajaran mesin yang berbeda untuk memodelkan teks yang diparafrasekan dengan teks lain. Tujuan

dari tulisan ini adalah untuk menganalisis kompleksitas permasalahan plagiarisme berdasarkan parafrase teks bahasa Indonesia dan membandingkan keunggulan metode klasifikasi dalam mendeteksi plagiarisme berdasarkan parafrase teks bahasa Indonesia. Dalam penelitian ini, data dari siswa dan 039; ringkasan tesis. Hasil penelitian ini untuk $K=4$ memberikan akurasi paling tinggi(V.A.R.Barao et al., 2022). Penelitian berikutnya menerapkan “Penerapan Algoritma Winnowing dan Word-Level Trigrams Untuk Mengidentifikasi Kesamaan Kata”. Perhitungan tingkat kesamaan berdasarkan kata (word-level) masih jarang dilakukan pada penelitian yang sebelumnya. Hasil pada penelitian ini bahwa algoritma Winnowing yang diterapkan menggunakan word-level trigrams dapat mendeteksi kesamaan pada teks sebesar 76,84%. Dari hasil penelitian dapat disimpulkan bahwa metode pencocokan pola dengan algoritma Winnowing dan word-level trigrams dapat digunakan untuk mengukur tingkat kesamaan teks(Ramdhani et al., 2022).

Pada penelitian sebelumnya yang dilakukan oleh (Saputra et al., 2022) dengan judul “Analisis Algoritma Winnowing pada Pendeteksian Plagiarisme Judul Tugas Akhir”, Tujuan artikel ini adalah menganalisis penggunaan algoritma winnowing dalam pengendalian plagiarisme judul tugas akhir. Proses pengendalian plagiarisme diawali dengan pembentukan rangkaian k-gram dan windows, tingkat kemiripan dihitung dengan koefisien Jaccard. Penelitian ini menggunakan value pair dan windows sebanyak 25 k-gram. Hasil penelitian menunjukkan bahwa variabel k mempunyai pengaruh yang lebih besar terhadap hasil kemiripan dibandingkan variabel w. Selisih hasil kemiripan pada nilai k yang berbeda adalah antara 5% hingga 9%, dan selisih hasil kemiripan pada nilai w yang berbeda adalah antara 1%

hingga 3%. Pada penelitian ini dapat disimpulkan bahwa penggunaan algoritma winnowing dengan nilai k yang berbeda dapat mempengaruhi tingkat kemiripan pada pengecekan plagiarisme. Dari hasil penelitian yang sebelumnya dapat diketahui bahwa algoritma Winnowing memiliki performa yang baik, pada penelitian yang dilakukan oleh (Mubarak, 2022) Penelitian ini hanya berfokus pada penerapan untuk pendeteksian plagiarisme pada file dokumen teks dan tidak memperhatikan sumber rujukan, ukuran dataset yang digunakan hanya menguji 10 dokumen teks, ukuran dataset yang digunakan terbatas sehingga tidak mencakup variasi yang cukup luas untuk menggeneralisasi hasilnya. Performa akurasi yang rendah, hasil pengujian menunjukkan bahwa tingkat akurasi terbesar yang dihasilkan adalah 47.58%, sedangkan tingkat akurasi terkecil hanya 19.28% dan tidak memberikan informasi yang lebih rinci tentang tingkat plagiarisme seperti tingkat persamaan antara sumber asli dan dokumen yang diuji.

Beberapa penelitian telah melakukan deteksi plagiarisme menggunakan algoritma fungsional yang sama dengan sidik jari dokumen, namun masalah sering muncul pada hasil k -gram, n -gram dan tingkat keakurasian berdasarkan nilai gram berbeda-beda.

Berdasarkan latar belakang tersebut, penelitian ini mengajukan judul tesis “Implementasi Sistem Deteksi Plagiarisme Berbasis Text” dengan tujuan untuk melakukan analisis performa yang baik dari hasil perbandingan algoritma cosine-similarity dan Jaccard-similarity pada algoritma pengenalan teks *winnowing* untuk mencari tingkat kemiripan antara teks. Sistem plagiarisme berbasis text fokus pada konten text yang ada di dalam dokumen yang ingin dibandingkan misalnya, dokumen text, artikel

web, buku elektronik, dan sebagainya. Dokumen meliputi informasi elektronik yang dibuat dengan memeriksa kesamaan antara berkas (file), diterima atau direkam dalam bentuk analog yang dapat dilihat, ditampilkan dan didengar oleh komputer atau sistem elektronik, namun tidak terbatas pada tulisan (Hermawan & Ismiati, 2020).

I.2 Rumusan Masalah

Rumusan masalah dalam penelitian ini adalah bagaimana perbandingan tingkat similarity antara algoritma Cosine Similarity dan Jaccard Similarity untuk deteksi kesamaan text terkait dengan variasi nilai N-gram.

I.3 Tujuan

Tujuan dari penelitian yang dilakukan adalah untuk memperoleh *performance* tingkat similarity algoritma Cosine Similarity dan Jaccard Similarity dalam deteksi kemiripan antara dua Text.

I.4 Manfaat

Berdasarkan uraian yang dijelaskan pada latar belakang masalah, maka manfaat penelitian ini diantaranya:

1. Untuk mengurangi dan menghindari adanya tindakan plagiarisme pada isi file dokumen karya ilmiah.
2. Untuk mengurangi resiko pelanggaran hak cipta.
3. Untuk lebih efisiensi dan penghematan waktu dalam proses penilaian

plagiarisme.

4. Untuk lebih memahami bagaimana cara pendeteksian kemiripan dua text.

I.5 Batasan Masalah

Batasan masalah dalam penelitian ini adalah data yang digunakan dalam pengujian dengan berbentuk text dan tidak mencakup hal lain seperti file gambar, video dan suara.

I.6 Sistematika Pembahasan

Dalam Sistematika penulisan penelitian tesis ini yang digunakan adalah sebagai berikut:

BAB I : PENDAHULUAN

Bab pendahuluan berisi gambaran umum penulisan terdiri dari latar belakang masalah penelitian dan merumuskan permasalahan yang akan diangkat dalam penelitian, selain itu tujuan di lakukannya penelitian, manfaat dan batasan masalah penelitian.

BAB II : TINJAUAN PUSTAKA

Bab tinjauan pustaka berisi tentang teori dasar yang menunjang dalam pembahasan penelitian yang dapat digunakan dalam menyelesaikan permasalahan yang berkaitan dengan penelitian.

BAB III : METODOLOGI PENELITIAN

Bab metode penelitian ini berisi tentang penjelasan mengenai perancangan penelitian diantaranya adalah metode pengumpulan data, metode yang diusulkan, data sekunder dan pengujian metode.

BAB IV : HASIL

Bab ini merupakan hasil analisis pengujian implementasi sistem program dari hasil algoritma yang diteliti.

BAB V : PEMBAHASAN

Bab ini merupakan hasil pembahasan dari hasil analisa metode tahapan-tahapan yang digunakan dalam penelitian.

BAB VI : KESIMPULAN DAN SARAN

Bab ini berisi kesimpulan dari hasil penelitian yang telah dilakukan dan juga saran bagi penelitian mendatang yang berasal dari kekurangan ataupun temuan penelitian lain.