

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan zaman di era industri 4.0 telah membuat media berita online menjadi sangat penting di kehidupan sehari-hari masyarakat untuk mencari sumber informasi. Hal ini berdampak pada wartawan media berita online untuk dapat mencari informasi berita yang cepat dan akurat, tidak menutup kemungkinan wartawan yang bekerja secara bersama-sama di lapangan melakukan tindakan plagiarisme ke wartawan lain atau mengambil bahan berita dari situs-media berita lain dan menggunakannya untuk dimuat dimediannya tanpa mencantumkan sumbernya [1].

Hal tersebut tentunya melanggar peraturan yang sudah ada karena melakukan tindakan plagiarisme, kasus plagiarisme sendiri dapat dideteksi dengan beberapa algoritma, seperti pada penelitian terdahulu yang menggunakan algoritma *winnowing* untuk mendeteksi plagiarisme judul tugas akhir mahasiswa, dengan 11 data uji didapatkan hasil rata-rata sebesar 66,30%, akan tetapi tanpa melalui proses *text preprocessing* dan hanya mendeteksi plagiarisme judul tugas akhir [2]. Pada penelitian yang menggunakan algoritma *cosine similarity* dan *TF-IDF* mampu mendeteksi kemiripan dokumen abstrak namun waktu pemrosesan lebih lama jika menggunakan *stemming* dan hasil *similarity* dari dokumen yang memiliki kemiripan tinggi akurasi hanya 0,56 dengan *stemming* dan 0,46 tanpa *stemming* dan jumlah kata rata-rata hanya 127-157 kata [3]. Pada penelitian yang membandingkan algoritma *jaro winkler* dan *latent semantic analysis* untuk deteksi

plagiarisme dokumen abstrak dengan 5 data uji didapatkan hasil rata-rata *similarity* algoritma *jaro winkler* sebesar 75,7% lebih baik dibandingkan dengan algoritma *latent semantic analysis* dan jika data yang dibandingkan sama persis algoritma *jaro winkler* mampu menghasilkan nilai plagiarisme sebesar 100%, sedangkan metode *latent semantic analysis* menghasilkan nilai plagiarisme sebesar 97,14% [4].

Oleh karena itu pada penelitian ini diusulkan algoritma *jaro winkler* untuk mengukur *similarity* dengan data berita online yang juga pernah dilakukan oleh peneliti lain untuk data yang berbeda [5]-[13].

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah disampaikan maka dirumuskan masalah yaitu, berapa hasil rata-rata *similarity* berita online yang diperoleh dari perhitungan algoritma *jaro winkler*.

1.3 Batasan Masalah

Adapun batasan masalah yang ada dalam penelitian ini adalah sebagai berikut :

1. Penelitian ini menggunakan data teks berita dari 20 situs media berita online daerah Kalimantan Tengah.
2. *Text preprocessing* yang dilakukan hanya *remove tag html, remove special character, case folding, number removal, filtering* dan *stemming*
3. Penelitian ini hanya dapat mendeteksi teks yang bersifat *plain* teks seperti huruf dan angka, tidak berupa gambar dan diagram.
4. Penelitian ini hanya mengukur *similarity* berita yang berbahasa Indonesia.

5. Penelitian ini berbasis website dengan menggunakan bahasa pemrograman PHP dengan framework Laravel.

1.4 Maksud dan Tujuan Penelitian

1.4.1 Maksud

Maksud dari penelitian ini adalah untuk mengetahui seberapa besar tingkat *similarity* antar berita online menggunakan algoritma *jaro winkler* serta mengevaluasi hasil dari *similarity* antar berita online.

1.4.2 Tujuan Penelitian

Tujuan dari penelitian ini adalah untuk mendapatkan nilai rata-rata hasil *similarity* berita online dengan menggunakan algoritma *jaro winkler*.

1.5 Manfaat Penelitian

Manfaat dari penelitian ini adalah untuk membantu memberikan referensi bagi para peneliti dalam pengembangan sistem atau aplikasi pengukuran *similarity* atau pengecekan plagiarisme khususnya berita dalam bahasa Indonesia.

1.6 Metode Penelitian

Metode awal yang digunakan adalah melakukan perancangan sistem yang akan dibangun dengan metode *prototype* dan setelah itu melakukan pengumpulan data. Data yang digunakan adalah data teks berita yang diambil dari 20 media berita online daerah Kalimantan Tengah. Langkah selanjutnya adalah proses *text preprocessing* yang mana data yang diperoleh akan melalui proses *text preprocessing* dengan tahapan *remove html tag*, *remove special character*, *case folding*, *number removal*, *filtering* dan *stemming*. setelah itu data yang sudah

melewati proses *text preprocessing* akan melewati proses implementasi pada algoritma *jaro winkler* untuk mencari nilai *similarity* dengan tahapan menghitung panjang string, menghitung nilai karakter yang sama, menghitung *transposisi*, menghitung jarak antara dua *string* dan menghitung nilai *similarity*. Langkah terakhir adalah proses evaluasi dari hasil *similarity* berita online yang dihasilkan dari perhitungan algoritma *jaro winkler*.

1.7 Sistematika Penulisan

Sistematika penulisan penelitian ini disusun untuk memberikan sebuah gambaran umum tentang apa yang ingin penulis teliti. Sistematika penulisan penelitian ini adalah sebagai berikut :

BAB I PENDAHULUAN

Bab ini menjelaskan tentang latar belakang permasalahan mengenai hal-hal yang menjadi kajian bagi penulis dalam melakukan penelitian, yang mana di dalamnya terdapat latar belakang, rumusan masalah, batasan masalah, maksud dan tujuan penelitian, manfaat penelitian, metode penelitian dan sistematika penulisan.

BAB II LANDASAN TEORI

Bab ini berisi kajian pustaka tentang penelitian lain yang telah dilakukan dan menjelaskan tentang teori-teori yang diperlukan untuk memahami permasalahan yang akan di bahas di penelitian ini, seperti halnya teori-teori tentang algoritma *jaro winkler* dan teori-teori yang berhubungan dengan topik penelitian.

BAB III METODE PENELITIAN

Bab ini menjelaskan tentang langkah-langkah pengerjaan dalam penelitian dan komponen-komponen yang digunakan untuk mencapai tujuan penelitian.

BAB IV HASIL DAN PEMBAHASAN

Bab ini berisi hasil pengujian algoritma *jaro winkler* yang telah dilakukan dalam penelitian, serta pembahasan terhadap hasil pengujian algoritma *jaro winkler*.

BAB V PENUTUP

Bab ini menjelaskan tentang kesimpulan yang diperoleh dari hasil implementasi dan pengujian, serta berisi saran untuk pengembangan terhadap penelitian selanjutnya.

