

ABSTRAK

Plagiarisme tindakan menjiplak atau mencuri karya tanpa pengakuan, menjadi tantangan serius dalam dunia akademis. Karya ilmiah sebagai target umum plagiarisme, semakin dipengaruhi oleh teknologi informasi. Penelitian ini mengimplementasikan sistem deteksi plagiarisme berbasis teks dengan membandingkan tingkat similarity antara algoritma Cosine Similarity dan Jaccard Similarity terhadap winnowing untuk deteksi kesamaan text terkait dengan variasi nilai N-gram 3, 5 dan 7. Pengujian dilakukan menggunakan bahasa pemrograman Python dan pustaka pendukungnya pada 20 dataset kalimat. Hasil pengujian menunjukkan bahwa Cosine Similarity lebih baik dalam mendeteksi kemiripan antara teks. Analisis akurasi menggunakan confusion matrix menghasilkan nilai akurasi sebesar 50%. Hasil perbandingan dari variasi n-gram yang berbeda-beda total performance similarity sebesar 15.89% dan rata-rata sebesar 0.26%. Sedangkan performance jaccard similarity total sebesar 13.59% dan rata-rata sebesar 0.23%. Meskipun Cosine Similarity memiliki akurasi yang lebih tinggi dibandingkan Jaccard Similarity, stabilitas tidak mencapai 100%.

Kata Kunci : Similarity Plagiarisme, Cosine Similarity, Jaccard Similarity, N-Gram, Data Teks