

ABSTRAK

Rajwan, 2026. **Komparasi Algoritma Winnowing Dengan Knn Dalam Mendeteksi Kemiripan Judul** (Dibimbing oleh **Fahrim Irhamna Rachman S.Kom.,MT** dan **Titin Wahyuni, S.Pd., MT**)

Peningkatan jumlah karya ilmiah mahasiswa di perguruan tinggi menyebabkan semakin tingginya potensi kemiripan judul penelitian. Kondisi ini dapat menimbulkan pengulangan topik, menurunkan orisinalitas karya ilmiah, serta menyulitkan proses verifikasi judul oleh pihak akademik. Penelitian ini bertujuan membandingkan kinerja algoritma Winnowing dan K-Nearest Neighbors (KNN) dalam mendeteksi kemiripan judul karya ilmiah mahasiswa. *Dataset* penelitian terdiri atas 956 judul skripsi mahasiswa Fakultas Ekonomi dan Bisnis Universitas Muhammadiyah Makassar. Tahapan penelitian meliputi *preprocessing* teks berupa *case folding*, tokenisasi, stopwords removal, dan stemming menggunakan Sastrawi. Algoritma Winnowing diterapkan melalui pembentukan k-gram, rolling hash, dan perhitungan Jaccard *Similarity*, sedangkan algoritma KNN menggunakan representasi TF-IDF dan cosine *similarity*.

Hasil penelitian menunjukkan bahwa algoritma KNN memiliki performa lebih baik dibandingkan Winnowing. KNN memperoleh rata-rata skor kemiripan sebesar 0,4327 dengan rata-rata waktu eksekusi 0,0006 detik, sedangkan Winnowing memperoleh rata-rata skor kemiripan 0,2100 dengan waktu eksekusi 0,0714 detik. Pada pengujian robustness, KNN juga lebih stabil dalam menangani variasi typo, perubahan urutan kata, dan penambahan kata. Kedua algoritma sama-sama mampu mendeteksi judul identik dengan skor sempurna 1,0000.

Berdasarkan hasil tersebut dapat disimpulkan bahwa KNN lebih efektif dan efisien untuk mendeteksi kemiripan judul secara kontekstual, sedangkan Winnowing lebih sesuai untuk mendeteksi kemiripan berbasis karakter literal. Penelitian ini memberikan kontribusi sebagai dasar pengembangan sistem deteksi kemiripan judul karya ilmiah yang dapat mendukung peningkatan kualitas pengelolaan akademik di perguruan tinggi.

Kata kunci: kemiripan judul, Winnowing, K-Nearest Neighbors, TF-IDF, cosine *similarity*.