

ABSTRAK

Plagiarisme dalam penulisan source code merupakan masalah serius di lingkungan akademik, di mana mahasiswa seringkali melakukan modifikasi kosmetik seperti mengubah nama variabel atau urutan fungsi untuk mengelabui sistem penilaian. Pendekatan deteksi berbasis teks mentah seringkali gagal mengidentifikasi kecurangan tersebut. Penelitian ini mengusulkan sistem deteksi kemiripan source code menggunakan kombinasi Abstract Syntax Tree (AST) untuk ekstraksi struktur logis, Term Frequency-Inverse Document Frequency (TF-IDF) untuk pembobotan token, dan Cosine Similarity untuk pengukuran kemiripan antar vektor. AST menormalisasi kode dengan mengabaikan elemen kosmetik, sementara TF-IDF memberikan bobot lebih pada elemen struktural yang langka. Hasil pengujian menunjukkan sistem mampu mendeteksi plagiarisme persis (verbatim), perubahan nama variabel (identifier renaming), dan modifikasi ringan dengan akurasi sangat tinggi (kemiripan di atas 97%). Melalui evaluasi 190 pasangan skenario pengujian, threshold optimal ditemukan pada nilai 0.65, menghasilkan akurasi sebesar 92.1% dan F1-Score 0.857. Sistem ini terbukti tangguh terhadap teknik obfuscation dasar, meskipun memiliki keterbatasan dalam mendeteksi perubahan paradigma pemrograman secara menyeluruh.

Kata Kunci: *Plagiarisme, Source Code, Abstract Syntax Tree, TF-IDF, Cosine Similarity.*