

ABSTRAK

Penelitian ini bertujuan membahas tentang analisis dan penerapan model deteksi kemiripan esai Bahasa Indonesia karya siswa SMA dan SMK dengan esai yang dihasilkan *ChatGPT* menggunakan model *IndoBERT*. Permasalahan utama yang diangkat adalah semakin maraknya penggunaan kecerdasan buatan (AI) dalam penulisan esai dapat mengancam kreativitas dan integritas akademik siswa. Penelitian ini bertujuan untuk menganalisis perbedaan esai asli siswa dan esai *ChatGPT* melalui kombinasi kemiripan semantik *IndoBERT* dan fitur linguistik, seperti keragaman leksikal, panjang kalimat, dan struktur sintaksis. Metode yang digunakan meliputi pengumpulan data esai siswa dan *ChatGPT*, persiapan data, ekstraksi fitur linguistik, pelatihan model *IndoBERT* untuk kesamaan semantik, dan klasifikasi menggunakan *neural network*. Hasil evaluasi menunjukkan bahwa model mencapai performa sangat tinggi dengan nilai AUC di atas 0,98 pada keempat model. Model tingkat paragraf tanpa pengetahuan *ChatGPT* menunjukkan hasil terbaik pada uji internal dengan *accuracy*, *precision*, *recall*, dan *F1-Score* yang semuanya mencapai 98%. Sementara itu, model tingkat kalimat terbaik diperoleh dari model dengan pengetahuan *ChatGPT* dengan *accuracy* 98%, *precision* 76%, *recall* 87%, dan *F1-Score* 80%. Fitur linguistik juga terbukti efektif sebagai variabel diskriminatif. Model yang dianalisis menunjukkan kemampuan generalisasi yang baik pada data di luar *dataset* utama dan diimplementasikan dalam aplikasi web berbasis *Flask* untuk tujuan analisis praktis di lingkungan pendidikan. Penelitian ini diharapkan dapat menjadi dasar untuk mengembangkan sistem deteksi teks AI yang akurat dan aplikatif untuk mendorong integritas akademis di Indonesia.

Kata Kunci: *ChatGPT*, deteksi AI, esai siswa, fitur linguistik, *IndoBERT*, kemiripan semantik