

ABSTRAK

Pesatnya perkembangan *Large Language Models* (LLM) seperti Llama 3.1 memungkinkan pembuatan teks berita sintesis yang sangat menyerupai gaya kepenulisan jurnalis manusia, sehingga menghadirkan ancaman disinformasi dan berita palsu di ruang digital. Dengan memanfaatkan karakteristik stilometri dan algoritma *Prior-Data Fitted Networks* (TabPFN), proses klasifikasi keaslian teks berita dapat dilakukan secara cepat, akurat, dan independen terhadap topik. Penelitian ini bertujuan membangun sistem klasifikasi teks berita sintesis AI dan berita asli manusia yang terintegrasi dengan metode *Explainable AI* (XAI) berbasis *Shapley Additive Explanations* (SHAP). Dataset yang digunakan berjumlah 500 dokumen berita berbahasa Indonesia dengan distribusi kelas seimbang, terdiri atas 250 artikel orisinal *Kompas.com* dan 250 teks sintesis Llama 3.1. Variabel input yang diekstraksi mencakup enam fitur stilometri tingkat permukaan, yaitu *Average Sentence Length* (ASL), *Average Word Length* (AWL), *Type-Token Ratio* (TTR), jumlah kata, jumlah kalimat, dan jumlah tanda baca. Proses pengolahan data mencakup sanitasi teknis tanpa menghilangkan tanda baca dan struktur morfologi kata, partisi data terstratifikasi dengan rasio 80:20 (400 data latih dan 100 data uji), serta inferensi *in-context learning* menggunakan TabPFN. Evaluasi kinerja menunjukkan bahwa model menghasilkan akurasi sebesar 89,0%, presisi 89,7%, *recall* 88,0%, dan *F1-score* 88,8%. Analisis keterbukaan model menggunakan SHAP membuktikan secara global bahwa *Type-Token Ratio* (TTR) merupakan fitur paling dominan dalam membedakan gaya kepenulisan, serta memvalidasi secara lokal sifat aditif kontribusi fitur dengan pergeseran probabilitas akhir menjadi 0,545 dari nilai dasar ekspektasi 0,485. Sistem ini diharapkan dapat menjadi instrumen forensik digital yang efektif dan transparan dalam mendukung verifikasi keaslian naskah berita daring.

Kata Kunci: Deteksi Berita AI, Stilometri, TabPFN, Explainable AI, SHAP.