

ABSTRAK

MUHAMMAD HASRADDIN HASNAN, Perkembangan teknologi informasi dan media sosial telah mempercepat penyebaran berita palsu, sehingga diperlukan sistem deteksi yang akurat, andal, dan mudah diinterpretasikan. Penelitian ini bertujuan mengembangkan sistem *deteksi fake news* berbahasa Indonesia dengan mengintegrasikan IndoBERT sebagai model klasifikasi teks dan SHAP sebagai pendekatan *Explainable Artificial Intelligence* (XAI) untuk menjelaskan kontribusi kata terhadap hasil prediksi. Dataset diperoleh dari TurnBackHoax dan Kaggle, kemudian melalui tahapan *preprocessing* berupa *cleaning* text, filtering bahasa Indonesia, tokenisasi, serta penyeimbangan data menggunakan *random oversampling* pada data latih. Dari 5.347 data awal, diperoleh 4.980 data setelah *filtering* bahasa Indonesia, terdiri atas 3.613 data valid dan 1.367 data *hoaks*. Data dibagi secara stratifikasi dengan rasio 80% untuk pelatihan dan 20% untuk pengujian. Setelah *oversampling*, data latih menjadi seimbang dengan masing-masing 2.890 sampel per kelas. Hasil eksperimen menunjukkan bahwa model baseline TF-IDF dan Logistic Regression memperoleh akurasi 77%, sedangkan IndoBERT mencapai akurasi 87%, dengan precision 0,87, recall 0,95, dan F1-score 0,91 pada kelas *hoaks*. Visualisasi SHAP menunjukkan token penting yang memengaruhi klasifikasi. Hasil ini membuktikan bahwa integrasi IndoBERT dan SHAP efektif meningkatkan deteksi berita palsu sekaligus memberikan transparansi model.

Kata Kunci: Deteksi Berita Hoaks, IndoBERT, SHAP, Explainable AI, Klasifikasi Teks.