
Klasifikasi Multi-label Artikel Ilmiah Berdasarkan Sustainable Development Goals Menggunakan Pre-trained Large Language Model

Adhistianita Safira Husna¹, Ade Romadhony², Suryo Adhi Wibowo³

^{1,2}Fakultas Informatika, Universitas Telkom, Bandung

³Fakultas Teknik Elektro, Universitas Telkom, Bandung

¹adhisti@students.telkomuniversity.ac.id,

²aderomadhony@telkomuniversity.ac.id,

³suryoadhiwibowo@telkomuniversity.ac.id

Abstrak

Tujuan Pembangunan Berkelanjutan (SDGs) adalah 17 agenda pembangunan global yang ditetapkan oleh Perserikatan Bangsa-Bangsa (PBB) melalui Kelompok Pembangunan Berkelanjutan PBB (UNSDG) untuk mencapai perdamaian dan kemakmuran bagi manusia dan planet ini sekarang dan di masa depan. Telah banyak karya ilmiah yang ditulis untuk mengatasi masalah terkait tujuan-tujuan ini. Namun, sistem klasifikasi artikel ilmiah saat ini terbatas pada bahasa Inggris. Sistem klasifikasi ini sangat penting karena menjembatani kesenjangan antara pengetahuan ilmiah Indonesia yang dihasilkan secara lokal dan kerangka kerja SDGs global, memungkinkan identifikasi, pemanfaatan, dan penguatan kontribusi penelitian Indonesia dalam mencapai pembangunan berkelanjutan di seluruh dunia. Penelitian ini mengembangkan sistem klasifikasi teks multi-label untuk mengklasifikasikan tesis dan artikel ilmiah Universitas Telkom ke dalam tujuan SDGs dalam bahasa Indonesia menggunakan metode transformer yang telah dilatih sebelumnya, dengan fokus pada model transfer learning bahasa tunggal (IndoBERT) dan multibahasa (mBERT, XLM-RoBERTa, dan mBART). Evaluasi dilakukan berdasarkan metrik berikut: F1-score, presisi, recall, akurasi subset, dan Hamming Loss; dengan dua pendekatan: word embedding yang telah dilatih sebelumnya dengan klasifikasi Multi-Layer Perceptron (MLP) dan klasifikasi fine-tuning end-to-end. Hasil eksperimen menunjukkan bahwa XLM-RoBERTa-large unggul sebagai model multibahasa dan mencapai kinerja optimal dengan F1-score 85,66% dalam pendekatan end-to-end, sementara model dasar juga mencatat F1-score 74,23% dalam pendekatan embedding kata yang telah dilatih sebelumnya dengan klasifikasi MLP. Hasil ini juga disebabkan oleh tantangan ketidakseimbangan dalam dataset. Oleh karena itu, hasil ini menunjukkan keefektifan model transformer dalam mengklasifikasikan teks multi-label dan potensi sistem untuk digunakan pada artikel ilmiah Indonesia.

Kata kunci: Fine-Tuning, Bahasa Indonesia, Klasifikasi Multi-label, Multi-layer Perceptron, Transformer
