KLASIFIKASI HALAMAN WEB JURNAL PREDATOR DENGAN DOC2VEC DAN *AUTOMATED MACHINE LEARNING* BERBASIS POHON DOM

Nama Mahasiswa : Ersan Karimi
NIM : 11201027
Dosen Pembimbing Utama : Gusti Ahmad Fanshuri Alfarisy, S.Kom., M.Kom.
Pembimbing Pendamping : Bowo Nugroho, S. Kom., M.Eng.

ABSTRAK

Jurnal predator mengancam integritas akademik karena menawarkan publikasi tanpa tinjauan sejawat yang layak. Berdasarkan studi Macháček and Srholec (2022), Indonesia menempati peringkat kedua global dengan 16,73% artikel yang diduga terbit di jurnal predator pada 2015–2017. Penelitian ini bertujuan untuk mengembangkan metode klasifikasi halaman web jurnal predator menggunakan kombinasi *Distributed Representations of Documents* (Doc2Vec) dan *Automated Machine Learning* (AutoML) berbasis struktur pohon *Document Object Model* (DOM). *Dataset* jurnal predator diambil dari situs *Kaggle*, sementara jurnal *non-predator* diambil dari *Directory of Open Access Journals* (DOAJ). Halaman utama situs jurnal dikumpulkan melalui web scraping dan dikonversi menjadi DOM corpus menggunakan dua pendekatan traversal, yaitu *Depth-First Search* (DFS) dan *Breadth-First Search* (BFS). DOM corpus kemudian diubah menjadi representasi vektor menggunakan Doc2Vec dan diklasifikasi secara otomatis dengan AutoML dari *Auto-Sklearn*. Evaluasi dilakukan dengan mengukur akurasi dan *macro avg F1-score* pada masing-masing kombinasi traversal dan waktu pelatihan AutoML. Pengujian dilakukan dalam rentang waktu pelatihan 15 hingga 120 menit, dengan interval 15 menit. Model terbaik pada traversal BFS diperoleh pada pelatihan 15 menit dengan macro avg F1-score sebesar 0.7812 dan akurasi 0.9196. Sementara itu, model terbaik pada data DFS diperoleh pada pelatihan 90 menit dengan macro avg F1-score sebesar 0.7853 dan akurasi 0.9255. Hasil ini menunjukkan bahwa metode traversal dalam pembentukan DOM corpus berpengaruh terhadap performa model klasifikasi jurnal predator. Traversal DFS cenderung menghasilkan performa lebih baik dibandingkan BFS dalam konteks kombinasi Doc2Vec dan AutoML berbasis pohon DOM, baik ditinjau dari nilai akurasi maupun *macro avg F1-score*.

Kata kunci: jurnal predator, klasifikasi, Doc2Vec, AutoML, DOM Tree