

ABSTRAK

Pendidikan merupakan komponen utama dalam membangun sumber daya manusia yang berkualitas. Ujian merupakan bagian dari proses evaluasi pendidikan untuk mengukur kemampuan siswa dalam memahami materi yang dipelajari. Proses ujian secara *online* memerlukan fasilitas mengenai pengelolaan soal, sehingga diperlukan klasifikasi untuk mengelompokkan soal sesuai dengan topiknya. Klasifikasi multi-label adalah proses pengelompokan data ke dalam beberapa kelas berdasarkan kesamaan ciri atau karakteristik data, di mana setiap soal dapat memiliki lebih dari satu topik. Penelitian ini berfokus pada pengklasifikasian soal mata pelajaran Bahasa Indonesia tingkat SMP dengan menggunakan metode *Problem Transformation* dan algoritma *Random Forest* serta *K-Nearest Neighbor* (K-NN). Metode *Problem Transformation* yang digunakan yaitu *Binary Relevance*, *Classifier Chain*, dan *Label Powerset*. Metrik evaluasi untuk menentukan kinerja terbaik yaitu berdasarkan *F1-Score* dengan *K-Fold Cross Validation*. Hasil penelitian menunjukkan bahwa algoritma *Random Forest* memberikan kinerja terbaik dibandingkan K-NN dengan nilai *F1-Score* terbaik di semua metode *Problem Transformation*. Nilai *F1-Score* terbaik dengan metode *Label Powerset* pada algoritma *Random Forest* sebesar 69%, dan K-NN sebesar 44%. Berdasarkan hasil tersebut, model klasifikasi soal dengan menggunakan algoritma *Random Forest* dan metode *Label Powerset* lebih efektif dalam mengklasifikasikan soal multi-label. Diharapkan penelitian ini dapat memberikan kontribusi dalam meningkatkan efisiensi pengelolaan soal ujian pada sistem pembelajaran *online* seperti *Learning Management System* (LMS).

Kata kunci— *klasifikasi multi-label, K-NN, machine learning, problem transformation, random forest*