

ABSTRAK

Al-Quran merupakan kitab suci yang mencakup berbagai bidang yang dapat menjadi pedoman umat Islam dalam kehidupannya. Al-Quran dapat dikategorikan sebagai teks multi-label untuk memudahkan umat Islam mempelajari Al-Quran. Penelitian tesis ini membahas klasifikasi teks dalam kasus dataset multi-label pada terjemahan Al-Quran bahasa Inggris dan berfokus pada penanganan masalah dataset yang tidak seimbang. Klasifikasi penelitian ini dibangun oleh metode yang diusulkan yaitu metode Ensemble dan Resampling. Hal ini dikarenakan keberhasilan penanganan dataset yang tidak seimbang pada studi sebelumnya, model tersebut dipilih. Meningkatkan performa yang buruk, adalah tujuan dari penelitian ini. Preprocessing data merupakan langkah pertama dalam penelitian ini, diikuti dengan resampling dan metode ensemble dalam klasifikasi sebagai kontribusi utama, dan terakhir adalah pengukuran prediksi klasifikasi. Cara kerja metode ansambel dalam klasifikasi adalah menggabungkan pembelajar berbasis-n menjadi pembelajar yang kuat. Skor F1 dan Kerugian Hamming digunakan dalam penelitian ini untuk mengevaluasi kinerja. Penelitian ini menggunakan F1-score dan Hamming Loss untuk melihat performanya. Berdasarkan beberapa percobaan, Metode yang diusulkan telah meningkatkan kinerja f1- score. Random Forest dapat meningkatkan f1-score sebesar 16%, sedangkan XGBoost dapat meningkatkan sebesar 14% dari studi sebelumnya. Di sisi lain, Resampling memberikan hasil yang berbeda tetapi tidak signifikan. Performa terbaik dari beberapa skenario tes dapat dihasilkan dari kombinasi Ensemble method, Resampling, penggunaan jumlah base learner, dan jumlah k-fold. Kombinasi ini mampu memberikan performa sebesar 50,12% pada f1- score dan 0,1002 pada hamming loss.

Kata kunci: Imbalanced, Ensemble, Bagging, Boosting, Multi-label, Classification, Resampling