

## ABSTRAK

Thesis ini membahas tentang penanganan imbalanced data pada klasifikasi teks multi-label dalam hadis bukhari. Penelitian ini menggunakan metode ensemble stacking karena dapat memberikan hasil evaluasi yang lebih baik pada kasus imbalanced data dibandingkan dengan klasifikasi dasar. Metode yang diusulkan umumnya menggabungkan beberapa prediksi dari metode klasifikasi dasar untuk mendapatkan hasil prediksi yang lebih baik. Dalam evaluasi, beberapa skenario eksperimen diterapkan untuk mengamati dampak kombinasi klasifikasi dasar serta jumlah fitur yang digunakan dalam klasifikasi. Selain itu, eksperimen dengan metode yang diusulkan juga dianalisis dibandingkan dengan baseline yang tidak menerapkan metode ensemble stacking dan membandingkan ensemble stacking dengan penelitian sebelumnya. Hasil percobaan menunjukkan bahwa kombinasi tiga pembelajaran dasar (Random Forest, Gaussian NB, dan SVM) dengan membatasi fitur yang digunakan hanya sebanyak dua ribu dapat memberikan hasil akurasi terbaik dibandingkan dengan eksperimen lainnya. Kombinasi pembelajaran dasar ini menghasilkan evaluasi akurasi 83.14 % dan F1-score 42.92 %.

**Kata kunci:** klasifikasi, multi-label, imbalanced data, ensemble stacking, akurasi, f1-score