



INTISARI

PENGARUH DEPENDENSI ANTAR *CLASSIFIER* PADA METODE *ENSEMBLE* DENGAN MODEL KLASIFIKASI BINER

Oleh

ARYA IMPUN DIAPARI LUBIS

23/526990/PPA/06669

Perkembangan *machine learning* mendorong penggunaan klasifikasi biner pada berbagai aplikasi, namun kinerja *classifier* tunggal sering tidak stabil ketika data kompleks, berisik, atau tidak seimbang. *Ensemble learning* banyak digunakan untuk meningkatkan performa, tetapi pengaruh dependensi antar *classifier* terhadap efektivitas penggabungan belum banyak dikaji secara khusus. Penelitian ini menganalisis pengaruh tingkat dependensi antar *classifier* terhadap kinerja *ensemble* linier berbasis *weighted average*. Evaluasi dilakukan dengan menetapkan *false positive rate* (FPR) sebesar 5% dan membandingkan *true positive rate* (TPR) pada tingkat FPR yang sama. Skor keluaran setiap *classifier* dikalibrasi menggunakan *Pool Adjacent Violators* (PAV) untuk memperoleh estimasi probabilitas posterior, ditransformasikan ke *log-likelihood ratio* (LLR), lalu digabungkan secara linier. Eksperimen sintesis digunakan untuk mengendalikan tingkat dependensi (melalui variasi korelasi), sedangkan pengujian pada data kontekstual menggunakan *Random Forest* dan *AdaBoost* serta pengukuran dependensi dengan *distance correlation*. Hasil menunjukkan bahwa penggabungan umumnya meningkatkan TPR pada dependensi rendah hingga sedang, namun manfaatnya berkurang dan dapat menjadi negatif pada dependensi sangat tinggi karena informasi antar *classifier* cenderung redundan. Pola pada data kontekstual sejalan dengan temuan sintesis, sehingga dependensi antar *classifier* perlu dipertimbangkan sebagai faktor penting dalam penerapan *ensemble* linier pada klasifikasi biner.



ABSTRACT

THE EFFECT OF CLASSIFIER DEPENDENCY ON ENSEMBLE METHODS IN BINARY CLASSIFICATION MODELS

By

ARYA IMPUN DIAPARI LUBIS

23/526990/PPA/06669

The rapid development of machine learning has increased the use of binary classification in many applications however, the performance of a single classifier is often unstable when the data are complex, noisy, or imbalanced. Ensemble learning is widely adopted to improve predictive performance, yet the effect of dependence between classifiers on the effectiveness of score combination has not been thoroughly investigated. This study analyzes how inter-classifier dependence influences the performance of a linear ensemble based on the weighted average. Model evaluation is conducted by fixing the false positive rate (FPR) at 5% and comparing the *true positive rate* (TPR) under the same FPR constraint. Each classifier output score is calibrated using the Pool Adjacent Violators (PAV) algorithm to obtain posterior probability estimates, transformed into log-likelihood ratios (LLRs), and then combined linearly. Synthetic experiments are used to control the dependence level (by varying correlation), while contextual experiments employ Random Forest and AdaBoost and quantify dependence using distance correlation. The results show that score combination generally improves TPR under low to moderate dependence, whereas the benefit diminishes and can become negative under very high dependence due to redundancy in the information carried by the classifiers. The contextual results are consistent with the synthetic findings, indicating that inter-classifier dependence should be considered an important structural factor when applying linear ensembles for binary classification.