

DETEKSI DINI PENYAKIT KANKER PARU DENGAN KOMBINASI METODE KLASIFIKASI *ADABOOST* DAN *RANDOM FOREST*

Roy Binsar Sinaga

Abstrak

Kanker paru merupakan jenis kanker yang paling banyak diderita kaum laki-laki serta terbanyak kelima di antara kanker lainnya pada kaum perempuan di Indonesia. Untuk mengetahui apakah seseorang menderita kanker paru atau tidak perlu dilakukan serangkaian tahapan diagnosis oleh tenaga medis yang cukup banyak dan tidak bisa didapatkan hasilnya secara cepat. Oleh karena itu pada penelitian ini, akan dilakukan percobaan untuk menggunakan kombinasi algoritma klasifikasi *Adaboost* dan *Random Forest* dalam mendeteksi dini kasus kanker paru berdasarkan serangkaian indikator atau atribut terkait kanker paru. Data yang digunakan dalam penelitian merupakan data sekunder yang berasal dari *website kaggle.com*. Data terdiri dari 309 *record* data dengan 10 fitur dan 1 kelas yang kemudian data diterapkan *preprocessing*. Dilanjutkan dengan pembentukan model yang terdiri dari dua model yaitu model *Random Forest* dan model *Adaboost* yang menerapkan *Random Forest* sebagai *weak learner*. Setelah model terbentuk dilakukan pengujian dengan data uji, dimana dilakukan pembagian data menggunakan *hold-out validation* sebanyak dua kali yaitu skala 70 : 30 dan 80 : 20. Hasil evaluasi menunjukkan bahwa kombinasi metode *Adaboost* dan *Random Forest* menghasilkan *accuracy*, *precision*, *recall*, dan *specifity* tertinggi masing-masing sebesar 95.40%, 96%, 96.30%, 96%. Hasil tersebut lebih baik dibanding dengan penerapan metode *Random Forest* tanpa *Adaboost* dimana menghasilkan *accuracy*, *precision*, *recall*, dan *specifity* tertinggi masing-masing sebesar 93.20%, 92.70%, 96.00%, 95.90%.

Kata kunci: Kanker Paru, *Random Forest*, *Adaptive Boosting*.

EARLY DETECTION OF LUNG CANCER WITH A COMBINATION OF THE ADABOOST AND RANDOM FOREST CLASSIFICATION METHODS

Roy Binsar Sinaga

Abstract

Lung cancer is the most common type of cancer in men and the fifth most common among other cancers in women in Indonesia. To find out whether a person has lung cancer or not, it is necessary to carry out a series of diagnostic stages by quite a lot of medical personnel and the results cannot be obtained quickly. Therefore, in this study, an experiment will be conducted to use a combination of the Adaboost and Random Forest classification algorithms in early detection of lung cancer cases based on a series of indicators or attributes related to lung cancer. The data used in this study is secondary data from the kaggle.com website. The data consists of 309 data records with 10 features and 1 class which is then preprocessed. Followed by the formation of a model consisting of two models, namely the Random Forest model and the Adaboost model which applies Random Forest as a weak learner. After the model is formed, testing is carried out with test data, where data is divided using hold-out validation twice, namely a scale of 70: 30 and 80: 20. The evaluation results show that the combination of the Adaboost and Random Forest methods produces accuracy, precision, recall, and specificity. the highest were 95.40%, 96%, 96.30%, 96%, respectively. These results are better than the application of the Random Forest method without Adaboost which produces the highest accuracy, precision, recall, and specificity of 93.20%, 92.70%, 96.00%, 95.90%, respectively.

Keywords: *Lung Cancer, Random Forest, Adaptive Boosting.*