



## ABSTRACT

An estimation of 627,000 women died of breast cancer in 2018 according to data published by the World Health Organization (WHO). Research on the detection of breast cancer is not a recent research topic in medical field, but this problem is considered insufficient to be a solution in the early detection of breast cancer. There are some researches on breast cancer detection using a machine learning approach. Some studies use ensemble methods to improve classification performance. One implementation of the ensemble is voting techniques. This technique combines several predictions from some of the constituent algorithms and conventionally uses the same weight. On the other hand, each algorithm has a different performance, so we need a method to determine the optimal weight of each algorithm in order to produce better classification performance compared to the voting method with balanced weighting. This study aims to develop a method for determining the optimal weight of each classifier making up the voting method. The classifiers that we use are Naïve Bayes (NB), Support Vector Machines (SVM), and Gradient Boosting (GB). The weighted voting method in this proposal is to determine the weight based on the ranking of the constituent classifier with the priority of accuracy, sensitivity, and specificity. This proposed method can achieve a sensitivity average score of 90.02% and an average specificity score of 96.25% and an average accuracy score of 93.92%. This research is expected to be used as a reference to develop weighting methods for the purpose of medical diagnosis as well as other fields related to the weighting method.

**Keywords:** Ensemble technique, classification, breast cancer, weighted voting



## INTISARI

Diperkirakan sekitar 627.000 wanita meninggal karena kanker payudara pada tahun 2018 menurut data yang diterbitkan oleh *World Health Organization* (WHO). Penelitian tentang deteksi adanya kanker payudara bukanlah topik yang baru dalam penelitian di bidang kesehatan, namun permasalahan ini dianggap masih belum cukup untuk menjadi solusi dalam deteksi dini kanker payudara. Terdapat banyak penelitian tentang deteksi kanker payudara menggunakan pendekatan *machine learning*. Beberapa penelitian menggunakan metode *ensemble* untuk meningkatkan kinerja klasifikasi. Salah satu implementasinya adalah dengan menggunakan teknik voting. Teknik ini menggabungkan beberapa prediksi dari beberapa algoritme penyusunnya dan secara konvensional menggunakan bobot yang sama. Di sisi lain, setiap algoritme mempunyai kinerja yang berbeda, sehingga diperlukan metode untuk menentukan bobot optimal dari masing-masing algoritme agar menghasilkan kinerja klasifikasi yang lebih baik dibandingkan dengan metode voting dengan pembobotan seimbang. Penelitian ini bertujuan untuk mengembangkan metode penentuan bobot optimal dari masing-masing *classifier* penyusun metode voting. Classifier penyusun voting yang digunakan yaitu *Naïve Bayes (NB)*, *Support Vector Machines (SVM)*, dan *Gradient Boosting (GB)*. Metode voting dengan pembobotan usulan yaitu dengan menentukan bobot berdasarkan peringkat *classifier* penyusun dengan prioritas akurasi, sensitivitas, dan spesifisitas. Metode usulan ini mampu mencapai rerata sensitivitas sebesar 90,02% dan rerata spesifisitas sebesar 96,25% serta rerata akurasi sebesar 93,92%. Penelitian ini diharapkan dapat digunakan sebagai acuan untuk mengembangkan metode penentuan bobot untuk tujuan diagnosis medis serta bidang lain yang berhubungan dengan metode pembobotan.

**Kata kunci** – Teknik ensemble, klasifikasi, kanker payudara, pembobotan voting