

ABSTRAK

Setiawan Junior (03082180037)

PERBANDINGAN METODE ALGORITMA *RANDOM FOREST* DENGAN *XGBOOST* DALAM MENDETEKSI KANKER PAYUDARA

(xvi + 53 halaman; 18 gambar; 3 tabel; 1 lampiran)

Penelitian ini bertujuan untuk menciptakan sebuah alat yang mampu digunakan untuk para kaum wanita dalam melakukan pemeriksaan secara mandiri dengan hasil yang jauh lebih akurat. Metode yang digunakan dalam memprediksi diagnosa kanker payudara adalah dengan metode *Random Forest* dan *XGBoost*. Data yang digunakan berasal dari data pasien di wilayah *Wisconsin* dengan total data 569 data mentah. Tahap perkembangan model machine learning ini adalah dengan membagi data mentah menjadi 75% *data training* sebesar 426 data dan 25% *data testing* sebesar 143 data. Data tersebut kemudian akan digunakan dalam membangun model machine learning, dimana akan membandingkan dua model algoritma *Random Forest* dan *XGBoost*. Hasil simulasi mendapatkan bahwa tingkat keakuratan *Random Forest* adalah sebesar 96% dan tingkat keakuratan *XGBoost* adalah sebesar 97%. Tingkat *recall Random Forest* adalah 0.96 dan Tingkat *recall XGBoost* adalah 0.94. Maka dapat disimpulkan bahwa model algoritma *Random Forest* lebih cocok dalam melakukan prediksi diagnosa kanker payudara.

Kata Kunci: *Machine Learning, Random Forest, XGBoost*

Daftar Pustaka: 17 (2021 – 2022)

ABSTRACT

Setiawan Junior (03082180037)

COMPARISON OF RANDOM FOREST ALGORITHM METHOD WITH XGBOOST IN DETECTING BREAST CANCER

(xvi + 53 pages; 18 figures; 3 tabel; 1 appendix)

This study aims to create a tool that can be used for women to conduct self-examinations with much more accurate results. Random Forest and XGBoost will be the method used in predicting the diagnosis of breast cancer. The data used is based from the patient data in Wisconsin with a total of 569 raw data. The development stage of this machine learning model is to divide the raw data into 75% training data with 426 data and 25% testing data with 143 data. The data will then be used to build a machine learning model which will compare the two models of the Random Forest and XGBoost algorithms. The simulation results show that the accuracy of Random Forest is 96% and the accuracy of XGBoost is 97%. The Random Forest recall is 0.96 and the XGBoost recall is 0.94. It can be concluded that the Random Forest algorithm model is more suitable in predicting breast cancer diagnosis.

Keyword: Machine Learning, Random Forest, XGBoost

References: 17 (2021 – 2022)