

ABSTRAK

Penyakit diabetes adalah salah satu masalah kesehatan masyarakat yang signifikan dan mempengaruhi jutaan orang di seluruh dunia. Penelitian ini akan melakukan analisis perbandingan dari tiga algoritma ensemble learning (Random Forest, AdaBoost, dan XGBoost) dalam mengklasifikasikan diagnosa penyakit diabetes. Berdasarkan penelitian yang telah dilaksanakan, disimpulkan bahwa model dengan akurasi tertinggi adalah Random Forest dengan nilai 0.86, setelahnya ada XGBoost dengan nilai 0.85 dan terakhir adalah AdaBoost dengan nilai 0.82. Dapat disimpulkan juga bahwa ketiga model memiliki performa yang baik dan dapat dipakai untuk melakukan klasifikasi penyakit diabetes. Berdasarkan visualisasi hasil *Feature Importance* yang telah dibuat, dapat disimpulkan bahwa algoritma *Random Forest* dan *XGBoost* memiliki kesamaan terhadap 3 fitur terpenting, yaitu Glukosa, BMI dan Umur. Sedangkan untuk *AdaBoost* 3 fitur terpenting adalah DPF, BMI dan Glukosa.

Kata Kunci – *Ensemble Learning, Data Science, Comparison, Feature Importance, Diabetes*

ABSTRACT

Diabetes is a significant public health problem and affects millions of people worldwide. This study will perform a comparative analysis of three ensemble learning algorithms (Random Forest, AdaBoost, and XGBoost) in classifying diabetes diagnoses. Based on the research that has been carried out, it is concluded that the model with the highest accuracy is Random Forest with a value of 0.86, XGBoost with a value of 0.85, and AdaBoost with a value of 0.82. It can also be concluded that the three models perform well and can be used to classify diabetes. Based on the visualization of the results of Feature Importance that has been made, it can be concluded that the Random Forest and XGBoost algorithms have in common the 3 most important features, namely Glucose, BMI and Age. As for AdaBoost, the 3 most important features are DPF, BMI and Glucose.

Keywords – Ensemble Learning, Data Science, Comparison, Feature Importance, Diabetes
