Combination of ADASYN-N and Random Forest in Predicting of Obesity Status in Indonesia: A Case Study of Indonesian Basic Health Research 2013

M Aqsha[1], SA Thamrin[1*], and Armin Lawi[2]

Abstract. Obesity is a pathological condition due to the accumulation of excessive fat needed for body functions. The risk factors for obesity are related to their obesity status. Various machine learning approaches are an alternative in predicting obesity status. However, in most cases, the available datasets are not sufficiently balanced in their data classes. The existence of data imbalances can cause the prediction results to be inaccurate. The purpose of this paper is to overcome the problem of data class imbalance and predict obesity status using the 2013 Indonesian Basic Health Research (RISKESDAS) data. Adaptive Synthetic Nominal (ADASYN-N) can be used to balance obesity status data. The balanced obesity status data is then predicted using one of the machine learning approaches, namely Random Forest. The results obtained show that through ADASYN-N with a balance level parameter of 1 ($\beta = 100\%$) after synthetic data generation and Random Forest with a tree number of 200 and involving 7 variables as risk factors, giving the results of the classification of obesity status which is good. This can be seen from the AUC value of 84.41%.