Abstract

Diabetes Mellitus (DM) is one of the dangerous diseases among adults and adolescents. The main cause of diabetes is increased levels of sugar in the blood as well as an unhealthy lifestyle is also one of the triggers of a person with diabetes. This occurs due to the disruption of insulin work that can not convert sugar levels into energy. According to the World Health Organization, there are over 400 million diabetics worldwide. With the large number of diabetics, a solution is needed to be able to detect diabetes disease quickly. This study aims to detect diabetics using naïve bayes algorithm by comparing missing value handling techniques, namely deleting records containing missing value and unsupervised imputation attributes. Naïve Bayes is one of the methods considered an effective algorithm used for the classification process. This study uses Medika's Sugar Dataset. The results showed by using the technique of filling missing values with mean and median values and using 3-fold cross validation get the best results with average accuracy of 74.9%, precision 90.64%, recall 66.57%, f-score 76.33%.

Keywords: diabetes, naïve bayes, classification, missing value, unsupervised imputation.